



A University of Sussex DPhil thesis

Available online via Sussex Research Online:

<http://sro.sussex.ac.uk/>

This thesis is protected by copyright which belongs to the author.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Please visit Sussex Research Online for more information and further details

C1, SAP and ZiCo:

Structural studies of three metal-binding proteins from a crystallographic perspective

Andrea Fallas

A thesis submitted for the degree of
Doctor of Philosophy (Biochemistry)
at the University of Sussex

9 September 2010

Statement

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.

Signature:.....

‘The woods are lovely, dark and deep,
But I have promises to keep,
And miles to go before I sleep,
And miles to go before I sleep.’

— *Robert Frost*

I dedicate this thesis to the memory of my friend Steven Keyes, who passed away during its completion due to complications of Fanconi Anemia.

<http://www.fanconi.org/>

<http://sites.google.com/site/stevensassociationmoonrisesite/sam>

Summary

Atomic resolution models of proteins are crucial for understanding their biological mechanisms and provide insights into the relationship between protein sequence and structure. Many protein structures incorporate metal ions, exploiting their unique chemistry as reaction centres or for structural stability. This thesis describes the progress made towards solving the structures of three such metal-binding proteins by means of X-ray crystallography.

Complement component C1 is a large protein complex that initiates the first line of immune defence and requires calcium for structural stability. Fragments of C1 have already been solved at high resolution, but there are no accurate models of the assembled complex. In this work, a new method for purifying intact C1 from human serum was developed and the purified complex was characterised by various methods. Finally, attempts were made to crystallise native human C1 with a view to obtaining high-resolution structures of the entire complex.

Serum amyloid P is another serum protein, also thought to be involved in the immune response. It is often found associated with amyloid deposits, although SAP binds a variety of ligands in a calcium-dependent manner. While the structure of SAP has been determined, its physiological function is still not fully understood. SAP was purified using established methods and its ligand-binding properties were investigated under various conditions using dynamic light scattering, in an attempt to gather more information about the possible function of this molecule.

Finally, ZiCo is a small peptide that was designed to switch between a multimeric coiled coil and a monomeric zinc finger fold on binding zinc. The system has been characterised extensively in solution, but high-resolution structures are required to validate the design. ZiCo was crystallised and diffraction data were collected. The structure of the peptide was partially solved, indicating that the multimeric form of the ZiCo peptide is indeed a trimeric coiled coil.

Acknowledgements

I would like to thank the following people for their contributions to my personal journey through this work. First and foremost, Darren Thompson. You gave me the opportunity to explore this fascinating subject in my own way and have been a fantastic teacher, mentor and friend throughout. I'll never be able to look at a conference room full of crystallographers in the same way ever again! My thanks also go to Sue Jones, for being a wonderful co-supervisor.

To Derek Woolfson and Eleonora Cerasoli, thank you for letting me cut my crystallographic teeth on ZiCo. I would also like to thank David Papapostolou and Beth Bromley for the help you have each given me with my work on this project. Louise Serpell, Karen Marshall, Tom Williams, Sarah Pannel and the other lab members, thank you for making the lab a fun and friendly place. I am deeply grateful for all your support during difficult times as well as day-to-day. I would also like to express my gratitude to Lucas Bowler and Julian Thorpe for your help with the mass spectrometry and TEM analyses, and Nigel Brisset for crystallography advice and tips. To Ray Francke and the other support staff, thank you for your excellent service.

To my university friends, especially Yalda Javadi, Caroline Norris and David Rose, thanks for all the good times and good advice. There is too much to say here, so I'll leave it at that. To my old friends, Pamela Heafield and Victoria Smith, thank you for your fantastic friendship and for your faith in me. And to my new friends, especially Jake Eley, Dan Stirling and the rest of the Brighton crew, your comradeship has been transformational. You have done more to help me finish this task than you will ever know. Stay safe.

Ed Inglis, thank you for giving me nowhere to hide.

And last, but certainly not least, Mama and Dad.

I would not be here without you. Thank you for everything. Danke.

Contents

Equations	vii
Tables	viii
Figures	x
Abbreviations	xiv
Chapter 1. Introduction	1
1. 1. Overview	2
1. 2. Proteins	3
1. 3. Metal-binding proteins	4
1. 4. Complement component C1	9
1.4.1 The complement system and C1	9
1.4.1.1 Complement activation pathways	10
1.4.1.2 C1-inhibitor	12
1.4.2 Structure and function of C1	14
1.4.3 Role in innate and adaptive immunity	17
1.4.3.1 Innate immunity and complement	17
1.4.3.2 Adaptive immunity and complement	18
1.4.4 Role in disease and therapy	19
1. 5. Serum amyloid P component	22
1.5.1 SAP and the pentraxins	22
1.5.2 Structure and function of SAP	24
1.5.3 Role in disease	27
1. 6. ZiCo: a designed switch peptide	28
1.6.1 Protein folding and design	28
1.6.2 Switch peptides	31
1.6.3 Design details	34
1. 7. Project background	37
1.7.1 C1: purification, characterisation and crystallisation	37
1.7.2 SAP: investigating metal and ligand binding	38
1.7.3 ZiCo: design validation	39
1. 8. Project aims	41
1.8.1 C1: purification, characterisation and crystallisation	41
1.8.2 SAP: ligand binding and co-crystallisation	41
1.8.3 ZiCo: design validation	42

Chapter 2. Background to methods	43
2. 1. Overview	44
2. 2. Protein purification	46
2.2.1 Precipitation	47
2.2.2 Chromatography	48
2.2.2.1 Affinity chromatography	48
2.2.2.2 Ion exchange chromatography	49
2.2.2.3 Size exclusion chromatography	50
2.2.3 Protein concentration	52
2. 3. Protein characterisation	53
2.3.1 Electrophoresis	54
2.3.1.1 Native PAGE	54
2.3.1.2 SDS-PAGE	55
2.3.1.3 Staining	56
2.3.2 Mass spectrometry	58
2.3.2.1 MALDI-TOF and peptide mass fingerprinting	59
2.3.2.2 Mass spectrometry of protein complexes	60
2.3.3 Dynamic light scattering	62
2.3.3.1 Scattering theory	62
2.3.3.2 Interpreting DLS data	63
2.3.3.3 Application to protein crystallisation	67
2.3.4 Circular dichroism spectroscopy	68
2.3.5 Transmission electron microscopy	70
2. 4. Protein crystallisation	73
2.4.1 Crystallisation theory	74
2.4.1.1 Precipitant effects	76
2.4.1.2 Protein crystal growth	80
2.4.2 Crystallisation methodology	83
2.4.2.1 Controlling nucleation and growth	83
2.4.2.2 Interpreting the results of a crystallisation experiment	84
2.4.2.3 Determining whether a crystal is protein or salt	87
2.4.2.4 Optimising crystallisation conditions	88
2. 5. X-ray crystallography	91
2.5.1 Principles of diffraction	92
2.5.1.1 Bragg's law and reciprocal space	94
2.5.1.2 Structure factors and Fourier transforms	96
2.5.2 The diffraction experiment	98
2.5.3 Cryocrystallography	100
2.5.4 Data collection and processing	102

2.5.4.1 General data processing procedure	102
2.5.4.2 Statistics from scaled and merged data	104
2.5.4.3 The Matthews coefficient	106
2.5.4.4 The Patterson function	107
2.5.4.5 The self rotation function	108
2.5.5 Phasing	109
2.5.5.1 Direct methods	109
2.5.5.2 Molecular replacement	113
2.5.5.3 Isomorphous replacement	116
2.5.6 Density modification	119
2.5.7 Model building and refinement	120
Chapter 3. Materials and methods	122
3. 1. Purification and characterisation of C1	123
3.1.1 Materials	123
3.1.2 Bioinformatics	123
3.1.3 Preparation of serum from human plasma	124
3.1.4 Precipitation of C1 from human serum	124
3.1.5 Purification of C1 by SEC	124
3.1.6 Further purification by IEX	125
3.1.7 Electrophoresis	126
3.1.7.1 Native agarose gel electrophoresis	126
3.1.7.2 SDS-PAGE	126
3.1.7.3 Staining	127
3.1.8 Mass spectrometry	127
3.1.9 Dynamic light scattering	129
3.1.10 TEM	129
3.1.11 Crystallisation	130
3. 2. Investigating the ligand binding properties of SAP	131
3.2.1 Materials	131
3.2.2 Purification of SAP	131
3.2.3 Aggregation assays	132
3. 3. Methods: ZiCo	133
3.3.1 Materials	133
3.3.2 ZiCo95	133
3.3.2.1 Crystallisation	133
3.3.2.2 SEC	134
3.3.3 ZiCo1	135
3.3.3.1 Crystallisation	135
3.3.3.2 CD	135
3.3.4 Crystallography of ZiCo1	136

3.3.4.1 Cryoprotectant screening	136
3.3.4.2 Cryocrystallography	137
3.3.4.3 Room temperature crystallography	137
3.3.4.4 MR models	137
3.3.4.5 Heavy atom soaks	138
Chapter 4. Purification and characterisation of C1	140
4. 1. Introduction	141
4. 2. Bioinformatic analysis	142
4. 3. Purification of native complement component C1	144
4.3.1 PEG precipitation	145
4.3.2 Purification by size exclusion chromatography	146
4.3.2.1 Choice of size exclusion matrix	146
4.3.2.2 Effect of buffers and pH on the purification of C1	149
4.3.2.3 Effect of ionic strength on the purification of C1	152
4.3.2.4 Choice of starting material: plasma or serum	154
4.3.2.5 Re-suspension volume	155
4.3.3 Summary: Purification by SEC	157
4. 4. Characterisation of C1 purified by SEC	158
4.4.1 Characterisation of SEC-purified C1 by electrophoresis	159
4.4.1.1 Native agarose gel electrophoresis	159
4.4.1.2 SDS-PAGE	162
4.4.2 Peptide mass fingerprinting of SEC-purified C1	165
4.4.3 Dynamic light scattering of SEC-purified C1	170
4.4.4 Transmission electron microscopy of SEC-purified C1	173
4.4.5 Storage of SEC-purified C1	176
4.4.6 Summary: Characterisation of C1 purified by SEC	177
4. 5. SEC-purified C1 under physiologically relevant conditions	178
4.5.1 The effect of ionic strength on SEC-purified C1	179
4.5.2 The effect of EDTA on SEC-purified C1	182
4.5.3 Divalent metal cations, pH and SEC-purified C1	184
4.5.3.1 SEC-purified C1 at physiological and acidic pH	185
4.5.3.2 Divalent metal cations at neutral and acidic pH	186
4.5.4 Melting point of SEC-purified C1	189
4. 6. Crystallisation of C1 purified by SEC	191
4.6.1 Pre-crystallisation test	191
4.6.2 Crystallisation kit for protein-protein complexes	194
4.6.3 Stura Footprint Screen	196
4.6.4 Structure Screen 1 & 2	197

4. 7. Further purification by ion exchange chromatography_____	200
4. 8. Summary: Purification by SEC and IEX_____	204
4. 9. Characterisation of Mono Q-purified C1_____	205
4.9.1 SDS-PAGE analysis of Mono Q-purified C1_____	205
4.9.2 Peptide mass fingerprinting of MonoQ-purified C1_____	207
4.9.3 Dynamic light scattering of Mono Q-purified C1_____	211
4.9.4 Storage of Mono Q-purified C1_____	214
4. 10. Mono Q-purified C1 under relevant conditions_____	215
4.10.1 The effect of ionic strength on Mono Q-purified C1_____	215
4.10.2 The effect of EDTA on Mono Q-purified C1_____	217
4.10.3 Calcium and EDTA at acidic and neutral pH_____	219
4.10.3.1 Mono Q-purified C1 at physiological and acidic pH_____	219
4.10.3.2 The effect of calcium on Mono Q-purified C1_____	220
4.10.4 Melting point of Mono Q-purified C1_____	224
4.10.5 The effect of Mono Q-purified C1 concentration_____	225
4. 11. Crystallisation of Mono Q-purified C1_____	226
4.11.1 Pre-crystallisation test_____	226
4.11.2 Structure Screen 1 & 2_____	226
4. 12. Discussion_____	229
Chapter 5. Investigating the ligand binding properties of SAP_____	237
5. 1. Introduction_____	238
5. 2. Purification of SAP_____	240
5. 3. Aggregation assays_____	243
5. 4. Discussion_____	246
Chapter 6. Validation of the ZiCo design_____	249
6. 1. Introduction_____	250
6. 2. Crystallisation of ZiCo1 and ZiCo95_____	251
6.2.1 ZiCo95 crystallisation attempts and analysis_____	252
6.2.1.1 Crystallisation trials_____	253
6.2.1.2 SEC analysis_____	255
6.2.2 ZiCo1 crystallisation and optimisation_____	258
6.2.2.1 Crystallisation trials_____	258
6.2.2.2 Optimisation_____	261

6. 3. Data collection and analysis	264
6.3.1 Data collection	264
6.3.2 Data processing	268
6.3.3 Twinning tests	270
6.3.3.1 Types of twinning	270
6.3.3.2 Implications for structure solution	271
6.3.3.3 Detecting merohedral twinning	272
6.3.3.4 SFCHECK	274
6. 4. Molecular replacement	278
6.4.1 Zinc finger model	278
6.4.2 Single helix models	280
6.4.3 Key residue models	284
6.4.4 Trimeric coiled coil models	287
6. 5. Automated model building and refinement	289
6. 6. <i>Ab initio</i> phasing	290
6. 7. Experimental phasing attempts	292
6. 8. Discussion	294
Chapter 7. Conclusions	301
7. 1. Conclusions	302
7.1.1 Complement component C1	302
7.1.2 Serum amyloid P component	303
7.1.3 ZiCo	304
7. 2. Further work	305
7. 3. Concluding remarks	306
Appendix I. Appendix I	319
i. C1 charge calculations	320
ii. Superose 6 calibration	321
iii. C1 crystallisation: SEC-purified C1	322
Appendix II. SFCHECK	325
i. ZiCo1 and ZiCo95 charge calculations	326
ii. Space groups	328
iii. Self rotation function	330

Equations

Equation 1:	Molar ellipticity	69
Equation 2:	Bragg's law	94
Equation 3:	Friedel's law	96
Equation 4:	The Matthews coefficient	106
Equation 5:	Structure factor equation	110
Equation 6:	Triplet phase relationship	110
Equation 7:	Relationship between structure factors of native and derivative crystals	117
Equation 8:	The crystallographic R-factor	121

Tables

Table 1:	Biophysical characteristics of Zico1	40
Table 2:	Scores used in assessing the results of crystallisation experiments	85
Table 3:	The seven crystal systems	93
Table 4:	Buffers used to re-suspend precipitated complement component C1	125
Table 5:	Gradients used in the purification of complement component C1 by IEX	126
Table 6:	Variables used in the PCT of ZiCo95	133
Table 7:	Buffers in the pH range 6.5–9 and their prevalence in the Biological Macromolecule Crystallisation Database	149
Table 8:	Comparison of two different gel preparation methods for peptide mass fingerprinting	166
Table 9:	Results of peptide mass fingerprinting on C1 purified by size exclusion chromatography	168
Table 10:	DLS measurements of SEC-purified C1 stored for 30 days under various conditions	176
Table 11:	The effect of ionic strength on the size and distribution of SEC-purified C1 by DLS	180
Table 12:	Reference ranges for common substances in human plasma	184
Table 13:	SEC-purified C1 at physiological and acidic pH	185
Table 14:	The effect of calcium, magnesium and zinc on SEC-purified C1	187
Table 15:	Results of the pre-crystallisation test on SEC-purified C1 at various protein concentrations	192
Table 16:	The number of conditions in which SEC-purified C1 was observed to precipitate at various proteins concentrations	194
Table 17:	Table of hits from Structure Screens 1 & 2 using SEC-purified C1	197
Table 18:	Complement-related proteins identified in Mono Q-purified C1	208
Table 19:	Dynamic light scattering analysis of the various fractions obtained by further purification of complement component C1 by IEX.	211
Table 20:	DLS analysis of Mono Q-purified C1	212
Table 21:	Effect of storing Mono Q-purified C1 at 4°C	214
Table 22:	The effect of ionic strength on the size and distribution of Mono Q-purified C1 by DLS	216
Table 23:	The effect of EDTA on Mono Q-purified C1	218
Table 24:	The size distribution profile of Mono Q-purified C1 at acidic and physiological pH	219
Table 25:	The effect of calcium and EDTA on Mono Q-purified C1 at pH 7.4	221
Table 26:	The effect of calcium and EDTA on Mono Q-purified C1 at pH 5.5	223
Table 27:	The size distribution of Mono Q-purified C1 at different protein concentrations	225
Table 28:	Structure Screen 1 & 2 crystallisation data for Mono Q-purified C1	227
Table 29:	Measurements derived from the DLS analysis of purified SAP	242
Table 30:	Promising conditions from initial ZiCo95 crystallisation screens	253

Table 31:	Initial hits from PEG conditions of the Stura Footprint screen	259
Table 32:	Initial hits from salt conditions of the Stura Footprint screen	260
Table 33:	Optimisation of ZiCo1 crystallisation conditions	261
Table 34:	Comparison of processing statistics in various space groups	268
Table 35:	Matthews' coefficient in various space groups	269
Table 36:	Native Patterson peak list	279
Table 37:	Data processing statistics for two ZiCo1 data sets	295

Figures

Figure 1.1:	Common metals in biological systems	6
Figure 1.2:	Overview of the activation pathways of the complement system	13
Figure 1.3:	The domain structure of C1q and association of subcomponents	15
Figure 1.4:	The proposed structure of the assembled C1 complex	16
Figure 1.5:	The pentameric structure of SAP	26
Figure 1.6:	The calcium binding site of SAP	26
Figure 1.7:	ZiCo design principles	33
Figure 1.8:	Classical Cys ₂ -His ₂ zinc finger	34
Figure 1.9:	The heptad repeat and packing of a trimeric coiled coil model	35
Figure 1.10:	Features of the ZiCo sequence showing the design principles	36
Figure 2.1:	Size distribution of 60 nm latex beads.	65
Figure 2.2:	Size distribution of a mixture of 60 nm and 220 nm latex beads.	65
Figure 2.3:	Size distribution by intensity, volume and number of a mixture of 60 nm and 220 nm latex beads	66
Figure 2.4:	Characteristic CD spectra of various secondary structure conformations of the polypeptide chain.	68
Figure 2.5:	Illustration of a typical protein phase diagram	75
Figure 2.6:	Illustration of a simple protein crystal in two dimensions	92
Figure 2.7:	Bragg diffraction	95
Figure 2.8:	The importance of phase information	97
Figure 2.9:	Diagrammatic representation of an X-ray diffraction experiment	99
Figure 2.10:	Distribution of atoms contributing to strong and weak reflections	111
Figure 2.11:	Three sets of lattice planes perpendicular to the zone axis	111
Figure 2.12:	The effect of introducing heavy atoms on the resulting structure factor	117
Figure 4.1:	Bioinformatic analysis of the complement component C1 sequence using Protparam	142
Figure 4.2:	C1 purification using Superdex 200	147
Figure 4.3:	C1 purification using Superose 6	148
Figure 4.4:	The effect of different buffers on C1 purified by SEC	150
Figure 4.5:	Effect of varying salt concentrations on the recovery of C1 from the precipitate	152
Figure 4.6:	Multiple extractions of PEG-precipitated proteins	156
Figure 4.7:	Native agarose gel electrophoresis of purified C1	160
Figure 4.8:	Comparison of PEG precipitated proteins and purified C1 under reducing and non-reducing conditions by SDS-PAGE	164
Figure 4.9:	SDS-PAGE analysis of native and activated C1 on 7% SDS-polyacrylamide gels	164
Figure 4.10:	Gel submitted for MS analysis of SEC-purified C1	166
Figure 4.11:	Size distribution of pooled fractions SEC-purified C1 by intensity	171

Figure 4.12:	Size distribution by intensity of individual fractions of the C1 peak purified by SEC	172
Figure 4.13:	Cross-linked C1 analysed by TEM	174
Figure 4.14:	Cross-linked C1 (for comparison)	175
Figure 4.15:	The effect of increasing concentrations of NaCl on SEC-purified C1	179
Figure 4.16:	The effect of EDTA on SEC-purified C1 by analytical SEC	182
Figure 4.17:	The effect of EDTA on SEC-purified C1 by DLS	183
Figure 4.18:	Size distribution of SEC-purified C1 in the presence of various metal ions at pH7.4	186
Figure 4.19:	Size distribution of SEC-purified C1 in the presence of calcium and copper at pH5.5	189
Figure 4.20:	Melting point determination of SEC-purified C1	190
Figure 4.21:	Crystals apparently growing from precipitated protein in the crystallisation drop	198
Figure 4.22:	Further purification of C1 using a Mono Q IEX column	201
Figure 4.23:	A modified further purification of C1 using a Mono Q IEX column	202
Figure 4.24:	Oxidised and reduced SDS-PAGE gels of the two major peaks eluting from a MonoQ separation of SEC-purified C1	206
Figure 4.25:	Venn diagram showing the distribution of unique, complement-related serum proteins between the two major peaks eluting from Mono Q purification of complement component C1	209
Figure 4.26:	Quantitative analysis of C1 subcomponents and C1 inhibitor identified in the two major peaks eluting from Mono Q purification of C1	209
Figure 4.27:	Size distribution by intensity of Mono Q-purified C1	212
Figure 4.28:	The effect of EDTA on the size distribution by intensity of Mono Q-purified C1	217
Figure 4.29:	The effect of EDTA on the size distribution by volume of Mono Q-purified C1	218
Figure 4.30:	Effect of calcium and EDTA on the size distribution by volume of Mono Q-purified C1 at pH7.4	220
Figure 4.31:	Effect of calcium and EDTA on the size distribution by volume of Mono Q-purified C1 at pH5.5	223
Figure 4.32:	Melting point determination of Mono Q-purified C1	224
Figure 4.33:	Aggregates and crystals in screens of Mono Q-purified C1	228
Figure 5.1:	SAP in various stages of purification	240
Figure 5.2:	Size distribution by intensity of purified SAP by DLS	242
Figure 5.3:	Aggregation of SAP monomers on the addition of calcium	244
Figure 5.4:	Adding PE to calcified SAP does not inhibit the further aggregation	245
Figure 5.5:	PE prevents calcium-induced SAP aggregation	245
Figure 6.1:	A comparison of the ZiCo1 and ZiCo95 sequences	250
Figure 6.2:	Elution profile of ZiCo95 in the presence and absence of equimolar zinc chloride	256
Figure 6.3:	Elution profile of reduced ZiCo95 in the presence and absence of equimolar zinc chloride	257

Figure 6.4:	Comparison of ZiCo1 batches in 50mM sodium phosphate, pH7.5, 50mM sodium chloride	263
Figure 6.5:	Diffraction pattern of ZiCo1 using 33% glycerol as a cryoprotectant	265
Figure 6.6:	Room temperature diffraction of ZiCo1	266
Figure 6.7:	Diffraction of ZiCo1 crystals grown in 1M citrate, 10mM borate, pH8.5, without any additional cryoprotectant	267
Figure 6.8:	Intensity distribution of P3 data	273
Figure 6.9:	Plot of the k^{th} moments of I for ZiCo data processed in space group P3	273
Figure 6.10:	SFCHECK output for the ZiCo data processed in P1	275
Figure 6.11:	SFCHECK output for the ZiCo data processed in P3	276
Figure 6.12:	SFCHECK output for the ZiCo data processed in P312	277
Figure 6.13:	Sequence alignment of ZiCo1 and 1GCM	280
Figure 6.14:	Detail of MR solution using a 24 residue polyalanine search model in P3	281
Figure 6.15:	Packing of MR solution using a 24 residue polyalanine search model	282
Figure 6.16:	ZiCo1 sequence with key residues highlighted	284
Figure 6.17:	Alignment of the ZiCo1 sequence with the sequence of a classical zinc finger	285
Figure 6.18:	MR solution using a 7-residue Phe10 model in P3	286
Figure 6.19:	MR solution for a 24-residue polyalanine trimeric coiled coil model in P1	288
Figure 6.20:	Model of ZiCo1 in P3 built using ARP/wARP	289
Figure 6.21:	Electron density maps calculated ab initio using ACORN	291
Figure 6.22:	A crystal of the ZiCo1 before and after being soaked in a heavy atom solution	292
Figure 6.23:	Packing of the refined ZiCo1 solution in space group P3	299
Figure 6.24:	Electron density near His6 in the refined ZiCo1 structure	300
Figure I.a:	Net charge of C1 as a function of pH	320
Figure I.b:	Superose 6 calibration curve	321
Figure I.c:	PEG MME 550 solubility curve	323
Figure I.d:	PEG 6000 solubility curve	323
Figure I.e:	MPD solubility curve	324
Figure I.f:	PEG 3500 solubility curve	324
Figure II.a:	The net charge of ZiCo1 as a function of pH	326
Figure II.b:	The net charge of ZiCo95 as a function of pH	327
Figure II.c:	Space group P1	328
Figure II.d:	Space group P3	329
Figure II.e:	Space group P312	329
Figure II.f:	Stereographic projection of the self rotation function in space group P1	330
Figure II.g:	Modified stereographic projection of the self rotation function in space group P1	331
Figure II.h:	Stereographic projection of the self rotation function in space group P3	332

Figure II.i:	Modified stereographic projection of the self rotation function in space group P3	333
Figure II.j:	Stereographic projection of the self rotation function in space group P312	334
Figure II.k:	Modified stereographic projection of the self rotation function in space group P312	335

Abbreviations

ATP	adenosine triphosphate
BMCD	Biological Macromolecular Crystallisation Database
BSA	bovine serum albumin
C1	complement component C1
C1-INH	C1 inhibitor
C4BP	C4b-binding protein
CD	circular dichroism
DLS	dynamic light scattering
DMF	dimethylformamide
EDAC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	ethylenediaminetetraacetic acid
IEF	isoelectric focusing
IEX	ion exchange chromatography
MAC	membrane attack complex
MALDI-TOF	matrix-assisted laser desorption/ionisation
MASP-1	MBL-associated serine protease 1
MASP-2	MBL-associated serine protease 2
MBL	mannan-binding lectin
MES	2-(N-morpholino)ethanesulfonic acid
MOβDG	methyl 4,6-O-(1-carboxyethylidene)- β -d-galactopyranoside
MOPS	3-(N-morpholino)propanesulfonic acid
MPD	2-methyl-2,4-pentanediol
MR	molecular replacement
MW	molecular weight
NPGB	4-nitrophenyl 4'-guanidinobenzoate
PDB	Protein Data Bank
pI	isoelectric point
PEG	polyethylene glycol
SAP	serum amyloid P
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SLE	systemic lupus erythematosus
TCA	trichloroacetic acid
TEM	transmission electron microscopy
Tris	tris(hydroxymethyl)aminomethane

Chapter 1

Introduction

1.1 Overview

This introductory chapter will set out the context of the present thesis. I will begin by introducing proteins and the importance of determining high-resolution models of their structures, focusing on the roles that both metals and metal-binding proteins play in the cause and treatment of disease. I will then describe the three different metal-binding proteins that are the topic of this thesis, in terms of both their structural features and our current understanding of their respective functions. I will underline the importance of unravelling the structures of these metal-binding proteins so as to deepen our understanding of fundamental biological processes — both for the sake of furthering basic scientific knowledge and because it offers the potential of modulating protein structures to treat disease. The background to my work will also be explored, and I will critically review the literature, as well as demonstrate how I hope to complement previous work on the three proteins under study. The chapter will conclude with a statement of the aims of my investigations into the structures of the serum proteins complement component C1, serum amyloid P component, and the designed switch peptide ZiCo.

1.2 Proteins

With the new millennium, we have entered the era of genomics. A rapidly growing number of genomes have been sequenced from a diverse set of organisms, and the information they contain is ready to be exploited. Aside from what genes can tell us about evolution and genetics, the power of genes lies in the proteins they encode. The word *protein* comes from the ancient Greek *prōtos*, meaning *first*; proteins are so called because they are a primary material of plants and animals (OED Online, 2008). Proteins have structural and mechanical, as well as catalytic functions. Fibrous keratins are the major component of the hair, scales, shells, silk and feathers that offer external protection to almost every species. Even viruses use protective protein coats to conceal their genetic material. The enzymes that catalyse the myriad of reactions inside cells are also almost exclusively proteins.

The key to the manifold functions of proteins lies in their three-dimensional structure, which is encoded by the sequence of amino acids that constitutes a protein. Precisely how proteins fold into their native state based on the information contained within their primary structure remains one of the great unanswered questions of biochemistry today. However, by using modern techniques to probe the molecular structure of proteins we can gain insights into their mechanisms to help further our understanding of biological processes and what happens when they go awry. Moreover, the design approach to drug discovery relies heavily upon high-resolution structures of disease-related protein targets and is becoming increasingly successful, with more and more designed drugs coming onto the market.

1.3 Metal-binding proteins

Metals are ubiquitous in biological systems and are essential for life. A variety of metals are employed as charge carriers, electron transporters, and catalysts in cellular processes throughout the plant and animal kingdoms. Photosynthesis, the process of converting light into chemical energy, is one of the most fundamental biological reactions. It requires a magnesium ion to be co-ordinated to the centre of a chlorophyll molecule, which absorbs and channels solar energy to drive the production of ATP, the ‘energy currency’ of the cell. Furthermore, ATP itself requires magnesium ions in order to be biologically active.

However, by far the most diverse role that metals play in biological systems is when they are found incorporated into proteins. This arrangement combines the function of the protein with the unique chemistry of the metal ion. For example, the alkali metals sodium and potassium carry charges across biological membranes, facilitated by protein pumps and channels, thereby regulating osmotic pressure and depolarisation. The alkaline earth metals calcium and magnesium both play major roles: calcium in stabilising protein structures, and magnesium in signal transduction pathways and catalysis. The transition metals, on the other hand, are exploited for their variable oxidation states and catalytic properties, which enable them to participate in a range of fundamental biological reactions.

Given their prevalence in biological systems and the vital role they play, it is not surprising that metals have long been used for medicinal purposes. The antibacterial properties of metals such as copper and mercury were well known to the ancient Greeks and Egyptians. In Ayurvedic medicine, which has its roots in the Indus valley, metals are central to the preparation of treatments for an assortment of ailments. Yet despite their widespread use in traditional medicine, metals as therapeutic tools in western medicine did not come of age until the early part of the twentieth century, when Paul Ehrlich used an arsenical drug for the specific treatment of syphilis.

Some decades later, Rosenberg et al., 1967, discovered that certain platinum compounds inhibited cell division. This led to the development of the revolutionary chemotherapeutic drug cisplatin, which is still widely used to treat a number of cancers by virtue of its ability to cross-link DNA and induce apoptosis of cancerous cells. Nowadays metals have become firmly established as therapeutic agents in modern medicine and their involvement in fundamental life processes is incontrovertible. Given their prevalence and the diversity of their functions, metal-binding proteins are of considerable interest to the medical and scientific community.

In order to understand proteins — and the metals they contain — at a molecular level, atomic resolution structural models are required. These are commonly obtained by nuclear magnetic resonance spectroscopy (NMR), which is confined to the study of small proteins up to around 30 kDa, and X-ray crystallography, which is primarily limited by the availability of protein crystals. For the purpose of this thesis, only crystallography will be discussed in detail.

Briefly, NMR exploits the magnetic properties of certain atomic nuclei to gain information about the chemical environment surrounding the nucleus. When placed in a magnetic field, nuclei that have spin — stable isotopes like ^1H with an odd number of protons and/or neutrons — absorb at characteristic frequencies. Variations in this resonant absorption arise depending on the local environment experienced by the nucleus (e.g. different functional groups), and are known as chemical shifts. These must be assigned to each atom in the NMR spectrum, and can be combined with restraints on distances and angles between atoms to give information about the structure and dynamics of an aqueous protein sample. For a comprehensive review of using NMR in structural genomics see Yee et al., 2006.

X-ray crystallography, on the other hand, uses the diffraction of monochromatic radiation by the electrons within a crystal lattice to deduce a map of the electron density contained within the unit cell of a protein crystal. A detailed overview of the technique is given in section 2.5, p. 91. By studying the detailed models that these two techniques provide and combining them with other methods to examine the solution

Biologically occurring metals are found clustered together in the periodic table where elements share common characteristics. The prevalence of these metals in life largely reflects their natural abundance and availability. Sodium and potassium form small, singly charged positive ions that are highly abundant in all living cells. They play important roles in maintaining osmotic pressure and regulating the electric potentials that drive such diverse processes as nerve impulses or stomatal opening. Similarly, the

small alkaline earth metals (calcium, magnesium) have roles in signal transduction and regulation, and many enzymes require these metals as cofactors. The transition metals are defined as elements whose atoms or cations have an incomplete d sub-shell. As a result of their partially filled d shells, transition metals possess a number of exclusive properties. They have the ability to adopt multiple oxidation states thanks to the relatively low reactivity of unpaired d electrons, which makes them ideal reactive centres in many enzymes. Transition metals often form coloured compounds, so the iron found in haemoglobin is responsible for the red colour of human blood while the blue colour of lobster blood is due to the copper-containing haemocyanin. Zinc is an essential biological trace element that is typically found as a divalent ion. It is relatively abundant and plays ubiquitous roles in the biochemistry of life, including a purely structural role in certain proteins (notably zinc-finger-containing transcription factors).

Metals can be used as drugs (platinum, for example, is used to therapeutic effect as described on page 5), but they can also be used as biochemical tools to study protein structures. A particular impasse in the undertaking of solving a protein structure by X-ray crystallography is the availability of phase information for the molecule being studied. This is required in order to transform the recorded diffraction data into a map of the electron density within the protein crystal. Phase information cannot be recorded directly; instead it must be derived using approaches such as molecular replacement, multi-wavelength anomalous diffraction and multiple isomorphous replacement. While molecular replacement simply borrows the phase information from a known homologous structure, the other two phasing methods rely on the properties of heavy metals to provide phases for the unknown structure directly.

Because of their increased number of electrons, heavy metals such as platinum, gold or mercury, scatter X-rays much more strongly than the lighter elements that constitute the bulk of macromolecular structures. If the contribution to the overall scattering is significant, the phases of the scattered X-rays will be similar to those which would be observed for the heavy atoms alone. The problem of solving protein phases is thus reduced to the problem of finding a few heavy atom positions in the unit

cell, which can be achieved using methods developed for small-molecule crystallography. Two common techniques for introducing heavy atoms into protein structures are to soak heavy metal compounds into protein crystals or to incorporate selenomethionine in the place of methionine by expressing proteins using specialised growth media for the host organism.

In summary, metals and metal-binding proteins are ubiquitous in biological systems and play an essential role in life as it is found on earth. The link between metals and disease has been established and the usefulness of metals as therapeutic agents has been explored. Finally, an outline has been given of the role of metals as tools in the study of protein structure. The next three sections of this chapter will describe three widely differing metal-binding proteins in terms of our current understanding of their respective structures and functions, with an emphasis on the importance of studying these proteins in order to deepen our understanding of both the fundamental biological processes they are involved in, and the role these proteins play in human health and disease.

1.4 Complement component C1

This next section will introduce a calcium-binding protein complex, complement component C1, which is a key player in the immune response. The complement system is an evolutionarily ancient branch of the immune system, in which C1 plays a decisive role. The state of our current understanding of the structure and function of the complex will be described, going on to discuss further the role that complement, and C1 in particular, plays in innate and adaptive immunity. This section will conclude by outlining the role of complement component C1 in disease and therapy.

1.4.1 The complement system and C1

The complement system is a subcomponent of the immune system, found in various forms throughout vertebrate and invertebrate species (Zarkadis et al., 2001). In humans, it comprises a host of serum proteins and receptors organised into a highly regulated and complex network. The system is primed for a cascade of activation, which produces a multitude of effectors that contribute to host immunity (Molina, 2004). Complement plays a crucial role in defending against infection and maintaining the integrity of an organism by targeting invading pathogens or aberrant cells. The nature of the complement system makes it a potent weapon in the defence against pathogens and other harmful entities. However, deregulation of this powerful system can have grave consequences for the host organism. Inappropriate activation of complement is thought to contribute to the development of various inflammatory diseases, and complement deficiencies are pathological too (Sjoholm et al., 2006). The role of complement in disease states is covered more fully in section 1.4.4, p. 19.

Complement activity was first described in 1888, but the term ‘complement’ itself was only coined at the turn of the century by Paul Ehrlich, to describe a heat

labile bactericidal substance present in normal serum that complemented the antibacterial action of antibodies in haemolytic systems (Skarnes and Watson, 1957). Today some thirty proteins that make up the biochemical cascade of complement have been identified and the role of complement in health and disease (as well as its therapeutic potential) are becoming increasingly clear (Wagner and Frank, 2010). Nonetheless, despite the increasingly large body of evidence that has been accumulated with a view to understanding the complement system, the precise details of its diverse pathways and functions are still largely speculative.

1.4.1.1 Complement activation pathways

Our present understanding is that there are three distinct mechanisms that result in complement activation, which converge towards the later stages of the reaction cascade. The most thoroughly investigated of these mechanisms is undoubtedly the classical pathway. The two other major pathways that activate complement are the mannan-binding lectin (MBL) pathway and the alternative pathway. An overview of the various pathways of complement activation is shown in Figure 1.2, p. 13.

The first component of the classical pathway is C1, a large protein complex that circulates in serum as a proenzyme. On recognising and binding to activators, it undergoes a conformational change, resulting in the activation of its serine protease subcomponents. Activated C1 subsequently cleaves the next protein in the complement cascade, C4. This cleavage generates a fragments that associates with C2 to form the catalytic subunit of C3 convertase. The C3 convertase in turn cleaves C3 to produce C3b, which binds to the surface on which complement was initiated. This process also generates C3a, which triggers a local inflammatory response. Cleavage of C3 precipitates the terminal events in the complement cascade: inflammation and

phagocyte recruitment; opsonisation of pathogens; and the assembly of the terminal complement components into the membrane attack complex (MAC), which ultimately results in the destruction of the target cell (Rother et al., 1998).

In addition to binding to pathogen surfaces, C3b also associates with the C3 convertase to give C5 convertase. Again, this complex cleaves the next protein in the cascade into C5a and C5b. The final reactions of the complement cascade involve the proteins C6 through C9, which associate with C5b. These reactions result in the formation of MACs on the surface of the target cell: pores in the cell membrane that disrupt cellular homeostasis and lead to destruction of the target cell.

While much research is currently being done on these pathways, the classical pathway remains the most studied to date. It has been connected with a number of diseases and is the most relevant to the present study. For this reason, the alternative and MBL pathways will not be discussed further here, but for comprehensive reviews of complement activation and its consequences see Duncan et al., 2008 and Sakamoto et al., 1998. Although C3 is clearly a key target for research into modulating the effects of complement, complement component C1 is interesting because it specifically controls activation by the classical pathway.

Furthermore, C1 shares significant structural and functional similarity with the activating molecule of the MBL pathway. MBL is analogous to C1q, while the mannose-associated serine proteases it joins with are analogous to the serine protease subcomponents C1r and C1s. In particular, MASP2 is very similar to the C1s molecule of the classical complement pathway and they are thought to have a common evolutionary ancestor (Stover et al., 2001). When the carbohydrate-recognising heads of MBL bind to specifically arranged mannose residues on the phospholipid bilayer of a pathogen, MASP2 is activated to cleave complement components C4 and C2 into C4a, C4b, C2a, and C2b. The structural and functional similarity between C1 and the MBL/MASP complex suggests

1.4.1.2 *C1-inhibitor*

C1-inhibitor (C1-INH) is a member of the serpin family of proteins, the archetypal serine protease inhibitors. Unlike most serpins, C1-INH possesses — in addition to its serpin domain — a unique N-terminal domain whose structure and function are as yet unknown (Wagenaar-Bos and Hack, 2006). Serpins share a common fold, which typically features three β -sheets and eight or nine α -helices as well as an exposed loop region known as the reactive centre loop (RCL). The RCL protrudes from the central β -sheet region and presents a pseudosubstrate to target proteases.

Serine proteases cleave proteins in a reaction that involves the formation of an ester bond between the enzyme and its substrate, giving rise to an acyl-enzyme intermediate. Subsequent hydrolysis of this ester bond usually regenerates the active site and releases the cleaved substrate. On cleaving the RCL of a serpin however, a rapid conformational change takes place — prior to hydrolysis of the acyl intermediate — in which the RCL inserts into a β -sheet at the distal end of the molecule, forming an additional β -strand. Since the protease is attached via an ester bond, this results in a loss of structural integrity and disruption of the active site, thereby stabilising the acyl intermediate and forming a covalent complex between the two proteins (Huntington et al., 2000).

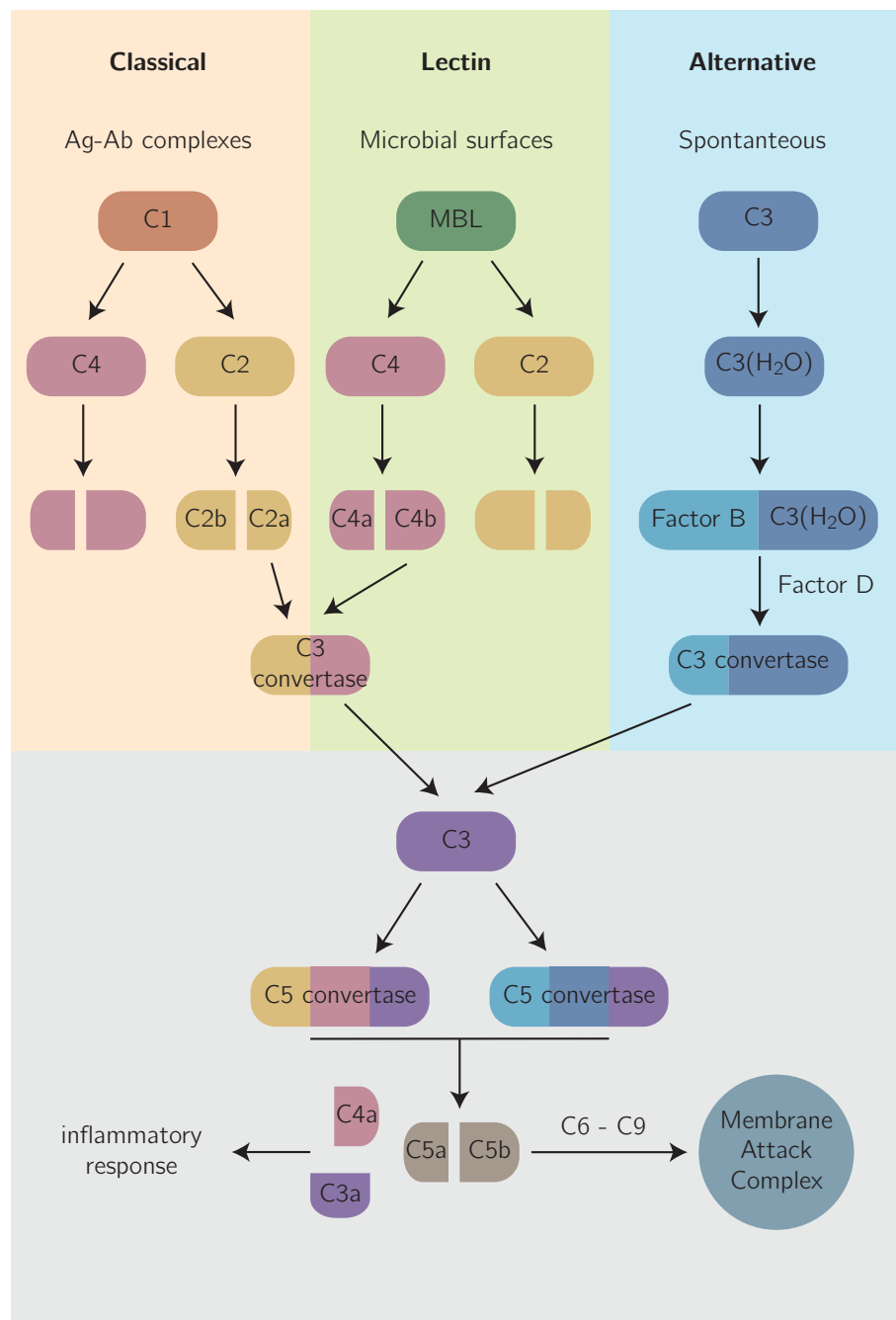


Figure 1.2: Overview of the activation pathways of the complement system. The most common activators are indicated for each pathway. The classical and lectin pathways are homologous, both generating the C4b2a complex (C3 convertase). The alternative pathway generates the C3bBb complex, a different C3 convertase. All three pathways converge at the level of C3 hydrolysis. The C3b fragment associates with either C3 convertase, transforming it into a C5 convertase. This cleaves C5, which precipitates the terminal events in the complex cascade, leading to formation of the MAC.

1.4.2 Structure and function of C1

Complement component C1 is an 800kDa complex composed of two functional subcomponents: a recognition element, C1q, and a catalytic core comprising two molecules each of the modular serine proteases C1r and C1s. The C1q subcomponent is itself composed of 18 homologous chains of approximately 25kDa — designated A, B and C — each possessing an N-terminal collagen-like domain and a C-terminal globular domain. In the absence of an activating signal, C1 exists as a proenzyme. Upon encountering a target molecule, C1q binds the ligand via the globular domains and relays the activation signal to the catalytic core via structural changes. These result first in the auto-activation of C1r, which subsequently cleaves C1s to yield the catalytic complex.

The first model for C1q was proposed by Reid and Porter, 1976, and this still forms the basis for our current understanding of the structure of C1q. The structure has six-fold symmetry and is commonly described as a bouquet of flowers: the collagen domains being the ‘stalks’ and the globular domains the ‘heads’. The individual chains are linked in the N-terminal region of their collagen-like domain by disulphide bonds to give three C-C dimers and six A-B dimers. The dimers then associate by forming a collagen-like triple helix to give three ABC-CBA structural units, which associate via strong non-covalent bonds to give C1q (Kishore and Reid, 2000). Furthermore, the C1q stalks are not straight, but splay outwards from the central fibrous portion. The collagen-like domains of the C1q chains contain a flexible hinge, resulting from a break in the continuity of the Gly-X-Y repeat sequence, which allows the ‘arms’ to diverge from the central stalk region (Schumaker et al., 1981). The accepted C1q structure just described is illustrated in Figure 1.3.

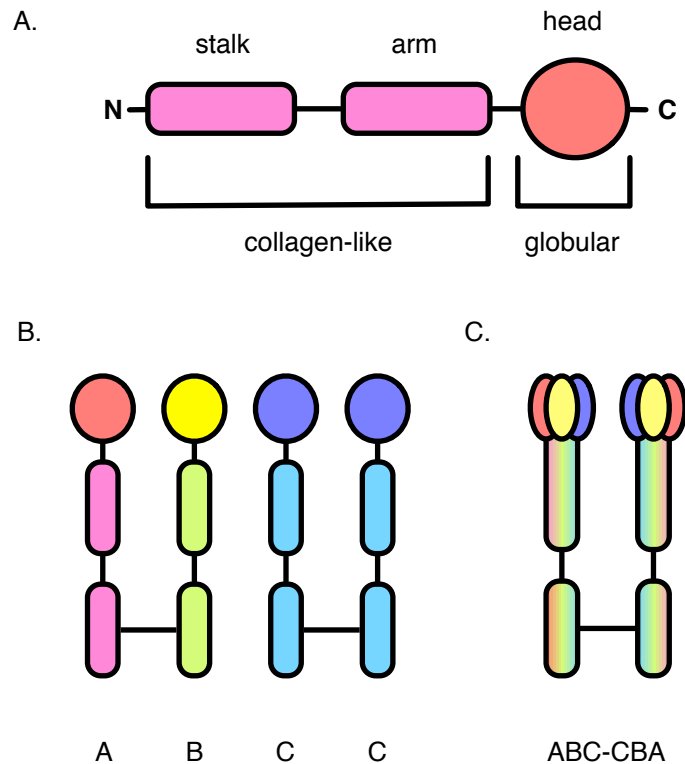


Figure 1.3: The domain structure of C1q and association of subcomponents. A. The domain structure of C1q showing the N-terminal collagen-like domain, which mediates subcomponent assembly; and the C-terminal globular domain, which is responsible for ligand binding. B. The association of A-B and C-C dimers via disulphide bonds in the N-terminal region. C. Two A-B dimers assemble with a C-C dimer in a triple helical collagen-like structure to give the ABC-CBA structural unit.

The catalytic core of C1 is a tetramer of the subunits C1r and C1s, which are homologous modular proteins. The C-terminal complement control protein modules (CCP1 and CCP2) mediate the activity of the serine protease (SP) domain, while the N-terminal CUB1 and EGF modules effect the assembly of the C1s-C1r-C1r-C1s tetramer (Villiers et al., 1985). This tetramer can be isolated independently of the C1 complex (Arlaud et al., 1980), and is thought to act as a scaffold for the attachment of C1q — and thus the assembly of the entire molecule — via the C1r moieties. Assembly of the core is dependent on the presence of calcium ions, which are bound with high affinity to C1r and C1s. In the accepted model of the tetramer, the catalytic domains of C1r are located at the centre and those of C1s at the ends of the assembly.

Within the C1 complex, the tetramer folds into a figure eight conformation around the arms of C1q. This brings the catalytic domains of C1r and C1s into contact, which is necessary for C1s activation by C1r (Arlaud et al., 2002). An illustration of the proposed structure for the assembled C1 complex is given in Figure 1.4.

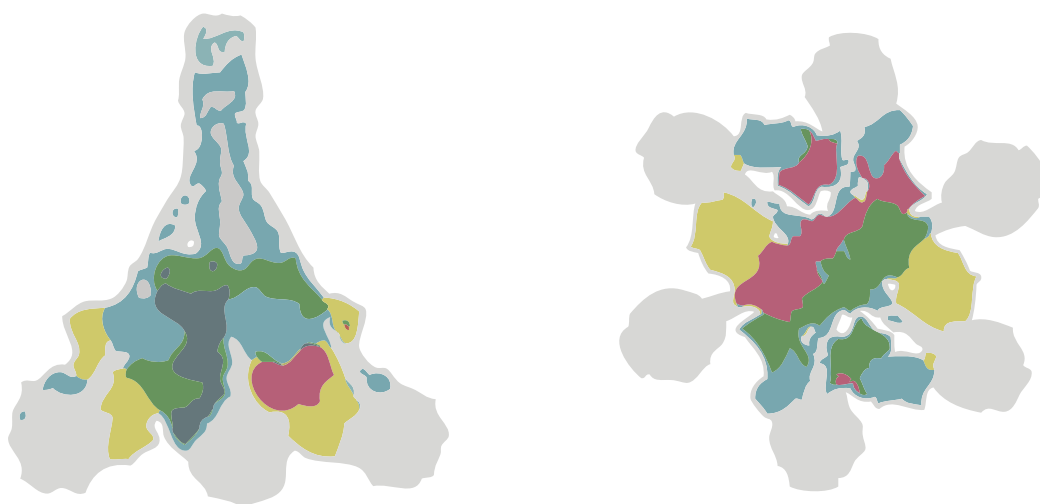


Figure 1.4: The proposed structure of the assembled C1 complex. The figure shows a schematic diagram based on the model derived by Gaboriaud et al., 2004, which builds on previously proposed theoretical models by incorporating high-resolution structural data from X-ray and NMR studies on fragments of the complex. The structure is shown as viewed from the side (left) and below (right). C1q is shown in grey, while the two C1r chains are shown in magenta and green. For C1s, the N-terminal serine protease domain is shown in yellow, while the C-terminal control domain is shown in blue. The figure shows how the C1q arms splay outwards from a central collagen-like stalk, while the serine protease subcomponents wrap around them, exposing the activated serine protease C1s.

1.4.3 Role in innate and adaptive immunity

Complement is established as a crucial player in innate immunity, the ancient branch of the immune system that provides an initial and non-specific barrier to infection. The innate immune system is phylogenetically much older than adaptive immunity. It is found in some form in all multicellular organisms, while the adaptive immune system is exclusive to all vertebrates except jawless fish. In lower vertebrates, complement is activated by the alternative and MBL pathways and its primary function is to promote the opsonisation of foreign material, while in higher vertebrates complement plays an important role as an effector and regulator of both types of immune response and is activated by all three pathways (Zarkadis et al., 2001). This suggests that the classical pathway of complement activation is a crucial component of the complex immune response specific to higher vertebrates and therefore particularly worthy of investigation in terms of human health and disease.

1.4.3.1 *Innate immunity and complement*

The human immune system can be viewed as being composed of distinct sequential reactions towards pathogens or aberrant cells. The complement system acts as an immediate barrier to infection, closely followed by the cell-mediated immune response, which involves the release of various cytokines and the action of phagocytes, natural killer cells and other leukocytes to rid the host of the offending antigen. Finally, after several days the adaptive immune response begins to exert its effects, mediated by T- and B-lymphocytes that act in response to antigen specific T-cell receptors and immunoglobulins or antibodies (Zipfel, 2009). This layered response caters for the short term need to neutralise the threat and at the same time offers the long term advantage of priming the immune system to recognise and remember previously encountered pathogens, generating immunity.

The role of the complement system in innate immunity has been reviewed fully elsewhere (Gasque, 2004), but the specific part that C1 and the classical pathway play in the crucial first stages of the immune response will be briefly outlined. As already mentioned, C1q binds to immune complexes (Cooper, 1985), toxic debris such as amyloid (Tacnet-Delorme et al., 2001), and epitopes on apoptotic cells (Paidassi et al., 2008). In addition to activating the complement cascade via its interaction with C1r and C1s, C1q independently stimulates myeloid cells via the C cell surface receptor CR1 and promotes phagocytosis of the foreign or toxic matter (Nicholson-Weller and Klickstein, 1999). The fact that individual complement components can elicit a variety of responses means that C1 exerts a level of control over a whole range of potential outcomes. Effectively, C1 triggers a self-contained system capable of destroying and removing pathogens, as well as serving to stimulate further immune responses in cellular immunity and the adaptive immune system.

1.4.3.2 Adaptive immunity and complement

While early experiments revealed a tentative link, it was the work of Mark Pepys that provided the first direct evidence linking complement to acquired immunity (Pepys, 1974). Complement is also now generally accepted to interact with the adaptive immune system, the lymphocyte-mediated defence system that confers long-term immunity against specific pathogens. The notion is now emerging that complement is not confined to particular roles in either innate or adaptive immunity, but acts as a bridge between both systems (Morgan et al., 2005).

Complement clearly plays an important role in defending an organism against attack by pathogens, but it is also important in maintaining homeostasis, for example by regulating tolerance to certain self-antigens (Prodeus et al., 1998) and promoting clearance of apoptotic cells (Ogden and Elkon, 2006). Given this dual role it is hardly surprising that complement has been linked with numerous diseases that are either the

result of inappropriate or excessive complement activation or caused by the absence of certain elements of an otherwise functional complement system. As a key activator of complement, C1 is uniquely placed in the wider context of the immune system, and certainly a protein worth further investigation in this respect.

1.4.4 Role in disease and therapy

Systemic lupus erythematosus (SLE) is a chronic autoimmune disease that can be debilitating or even fatal. While it can be treated (mainly with steroids and immunosuppressant drugs) it remains incurable. Deficiencies of any of the proteins involved in the classical pathway, particularly in the early stages of complement activation, are associated with a predisposition to SLE, and complement itself is implicated in the characteristic inflammatory autoimmune response (Walport, 2002).

Complement activation has been implicated in pulmonary disorders such as asthma (Regal, 1997), as well as cardiac disorders such as atherosclerosis (Niculescu and Rus, 1999) and tissue injury following myocardial infarction (de Zwaan et al., 2003). Atherosclerotic heart disease is a leading cause of death in western countries and is thus of considerable interest to the research and medical community. The current consensus is that the disease is a consequence of chronic inflammation of the coronary arteries. Over time, the normally protective inflammatory response becomes part of the disease process itself (Ross, 1999). Complement activation in atherosclerosis is thought to occur via both the alternative and classical activation pathways (Niculescu and Rus, 1999). The function and regulation of the complement system in cardiovascular disease is broad and complex, and will not be discussed further here. However, the reader is directed towards a recent review of the subject for further information (Oksjoki et al., 2007).

Perhaps unsurprisingly, given its role in clearing foreign bodies and damaged cells, complement is also involved in other types of tissue damage, for example

traumatic brain injury (Stahel et al., 1998) and rejection of transplanted tissues and organs (Sacks et al., 2003). Indeed, complement activation via the classical pathway is thought to contribute to pathogenesis in conditions as diverse as Alzheimer's disease (D'Andrea, 2005) and HIV infection (Stoiber et al., 2003). For a comprehensive review of the therapeutic applications of the complement system, see Wagner and Frank, 2010. Complement component C1 thus constitutes a very attractive target for developing drugs to inhibit undesired activation of the system. Aside from the benefits these could bring to patients, successful treatment with a complement inhibitor is the only conclusive way to prove a link between complement activation and a given disease, since the presence of elevated levels of complement components does not necessarily constitute a causative effect.

There are several approaches to inhibiting complement. The first is to block the pathway at the central stages with the aim of extending the inhibitory action as widely as possible. The problem with this approach is that it does not take into account either the diverse mechanisms of complement activation and their contributions to pathophysiology, or their protective effects (Mollnes and Kirschfink, 2006). A far more effective approach would be to target specifically those proteins involved in the initial stages of the various activation pathways, since this would allow the development of precise therapeutic tools with which to treat and investigate complement-related diseases.

Naturally, the development of inhibitory compounds for clinical applications is generally greatly aided by the availability of detailed structural information about target molecules and an understanding of the processes they are involved in. It follows that C1, as the first component of the classical pathway and thus a key player in numerous diseases, represents a particularly attractive subject for further investigation. In fact, the naturally occurring C1-inhibitor has been used in a clinical setting to treat angioedema, sepsis and myocardial infarction (Clemenza et al., 2003).

Widespread therapeutical application of C1-inhibitor is nonetheless hampered by difficulties in producing large quantities of C1-INH: isolating the protein form

donated blood is expensive and risky, while recombinant C1-INH does not possess the same glycosylation as the natural human form (Koles et al., 2004). A full understanding of the structure and action of complement component C1 would thus appear to be of great interest, both from a therapeutic perspective as well as for the sake of basic science.

1.5 Serum amyloid P component

This section will introduce serum amyloid P component (SAP), another protein involved in immunity and associated in particular with Alzheimer's disease. SAP is a member of the pentraxin family of proteins, and this group of immune modulators will be described with a particular focus on the structure and function of SAP. I will then outline our current understanding of the role of SAP in disease, and describe the potential applications of my work in the wider scope of the life sciences.

1.5.1 SAP and the pentraxins

SAP is a glycoprotein found in normal plasma. It is a member of the pentraxin family, which also includes the acute phase protein, C-reactive protein (CRP). The pentraxins are a highly conserved family, characterised by their calcium-dependent ability to bind a variety of ligands. Functionally entwined with the complement proteins described in the previous section, pentraxins are also involved in the innate immune response (Bottazzi et al., 2006).

Pentraxin monomers exhibit an overall β -jelly roll topology, similar to the legume lectins, but their name derives from their distinctive ultra-structure. Five identical subunits associate non-covalently to form a ring with five-fold symmetry, approximately 10 nm in diameter. Members of the pentraxin family have been found to be highly conserved among vertebrates and even show homology to proteins from the distantly related horseshoe crab, *Limulus polyphemus*. This suggests that they provide a distinct evolutionary advantage, implying that they play an important physiological role (Gewurz et al., 1995).

As components of humoral immunity, the pentraxins can be grouped with members of the complement cascade, collectins and ficolins under the heading of soluble

pattern recognition receptors. These molecules are key effectors and modulators of immunity, and represent the functional ancestors of antibodies (Mantovani et al., 2008). The family can be split into two distinct groups, based on sequence characteristics. The short pentraxins are typified by CRP and SAP, which are produced in the liver, while the classic long pentraxin is PTX3, which is secreted by a variety of different cell types. The short pentraxins exert their effects at a systemic level, while the effects of PTX3 are locally targeted.

CRP was the first pentraxin to be identified, as a serum component which precipitated the C-polysaccharide produced by *streptococcus pneumoniae* during acute infection (Tillett and Francis, 1930). Since its discovery, much work has been done to elucidate the structure and function of CRP, and it is now used a routine diagnostic aid in the treatment of heart disease. As an acute phase protein, levels of CRP rise dramatically during the inflammatory response, making it a useful marker for inflammation. Serum levels can be easily measured using ELISA for example, and can be used to chart the progress of a disease or the effectiveness of treatments.

However, despite its widespread clinical use, the biological function of CRP has not been unequivocally defined. This is in part due to the considerable differences between species in terms of sequence, and above all in the regulation of CRP. For example, while CRP is the classical human acute phase protein, it is not an acute phase protein in the mouse. Conversely, a major acute phase reactant in mouse is SAP, which demonstrates no comparable function in man. This has limited the use of traditional genetic approaches to unravelling the physiological role of both CRP and SAP. In contrast, studies of the long pentraxin PTX3 using genetically modified mice have clearly defined the role of this molecule at the crossroads of innate immunity, matrix assembly, and female fertility (Bottazzi et al., 2006).

1.5.2 Structure and function of SAP

The major known physiological function of SAP is its binding to various substances *in vivo*, including DNA, apoptotic cells and — of course — amyloid (Pepys and Butler, 1987; Bijl et al., 2003; Pepys et al., 1979). Trimers of human SAP have been shown to activate classical complement by binding strongly to C1q (Ying et al., 1993), and both SAP and CRP opsonise apoptotic cells for phagocytosis by interacting with Fc γ receptors (Mold et al., 2002). The contribution of SAP to maintaining immunological tolerance to self is demonstrated by the observation that SAP knockout mice, in which the gene encoding the protein has been deleted, spontaneously develop an autoimmune disease similar to human SLE (Bickerstaff et al., 1999). However, targeted depletion of the protein by a competitive inhibitor of amyloid binding, which causes the SAP molecules become cross-linked dimers, was developed as an approach to treating diseases such as systemic amyloidosis, Alzheimer’s disease and type 2 diabetes (Pepys et al., 2002). SAP inhibition by this method shows no negative effects in toxicological studies.

In normal human serum, SAP exists as a pentamer of identical 25 kDa subunits that form a non-covalent glycosylated protein complex (Hutchinson et al., 2000). The pentameric structure of SAP was solved over a decade ago (Emsley et al., 1994), closely followed by the structure of CRP bound to phosphocholine (Thompson et al., 1999). These seminal studies have furthered our understanding of the structure and function of the pentraxins. A crystal structure of SAP in complex with the cyclic pyruvate acetal of galactose has since been described by Thompson et al., 2002, shedding yet more light on some of the ligand-binding properties of SAP.

Although SAP exists as a pentamer *in vivo*, in the absence of calcium it forms stable decamers composed of two pentamer rings interacting face to face (Pepys et al., 1997). In contrast, in the presence of calcium, isolated human SAP auto-aggregates and precipitates from solution. This calcium-dependent aggregation is inhibited by all

known SAP ligands that require calcium for recognition, by physiological concentrations of albumin, and by elevated salt concentrations. The molecular basis for this behaviour can be explained in terms of the crystal structure of SAP (Thompson et al., 2002).

The native pentameric SAP assembly is shaped like a donut, approximately 100 Å in diameter and 35 Å deep. It possesses C5 symmetry, with the subunits arranged like the petals of a flower around the rotational symmetry axis (see Figure 1.5). Each subunit contains two calcium binding sites. Calcium I is co-ordinated by two ligands on Asp-58, and single ligands on Asn-59, Glu-136, Gln-137 and Asp-138. An additional seventh ligand is provided by a buffer ion. Calcium II is more loosely bound to the protein and is co-ordinated by the side-chains of Glu-136, Asp-138, Gln-148 and two water molecules (see Figure 1.6).

Crystals of SAP can be grown at pH5.5 in the presence of high calcium and acetate ion concentrations, conditions that circumvent the tendency towards excessive aggregation in the presence of calcium ions at near neutral pH. Interestingly, it has been shown that the calcium-dependent and -independent binding of SAP to glycosaminoglycans and amyloid is slightly enhanced at acidic pH (Danielsen et al., 1997). This has relevance in a biological context, since the pH is typically lowered at sites of infection or inflammation, lending weight to the validity of SAP structures from crystals obtained under these conditions.

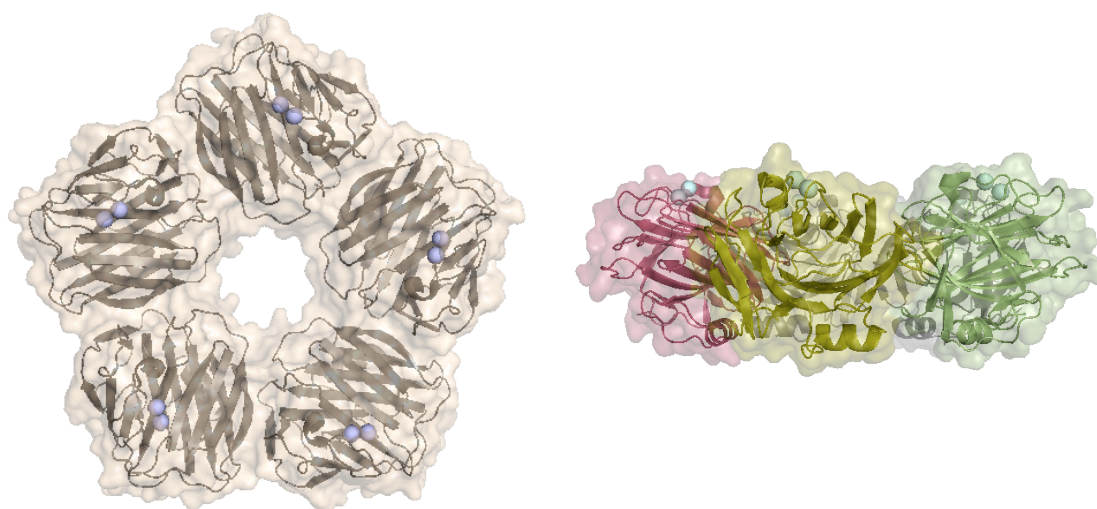


Figure 1.5: The pentameric structure of SAP. The top view (left) shows the five-fold symmetry of the assembly, with the subunits arranged in a ring around the symmetry axis. The calcium ions are shown in blue. The side view (right) shows the calcium ions bound to one face of the pentamer, with each of the subunits a different colour.

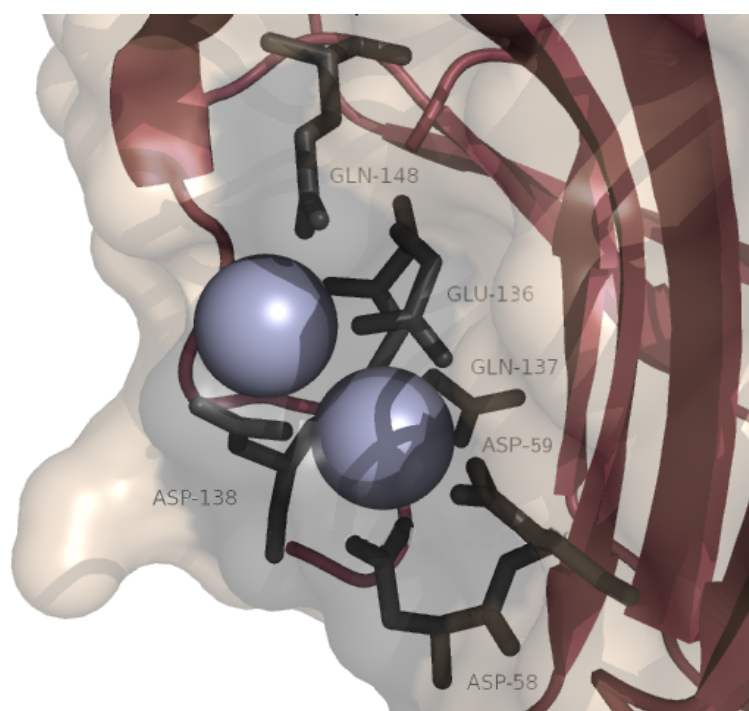


Figure 1.6: The calcium binding site of SAP. The calcium I ion (right) is singly co-ordinated by Asp-59, Glu-136, Gln-137 and Asp-138, and doubly co-ordinated by Asp-58. An additional ligand is provided by a buffer molecule (not shown). The calcium II ion (left) is more loosely bound to SAP, and is co-ordinated by Glu-136, Asp-138, Gln-148 and two water molecules.

1.5.3 Role in disease

The precise physiological function of SAP is still unclear and experimental evidence with regard to the role of SAP in disease has yielded little in the way of firm answers. SAP is known to bind a variety of ligands — most notably amyloid — and this is likely to be the key to its biological role. Like its homologue CRP, it is able to activate the classical pathway of complement (Hicks et al., 1992), although others have speculated that SAP is primarily involved in regulating complement activation (Garcia de Frutos and Dahlback, 1994). The interplay between the complement system and SAP places it squarely in the context of numerous diseases that have an inflammatory component.

Alzheimer's disease is perhaps the most widely known of the diseases involving amyloid but numerous other conditions also feature amyloid deposits. Among these are various genetic disorders but also inflammatory diseases such as rheumatoid arthritis and atherosclerosis. As an amyloid-binding protein linked with inflammatory immune responses SAP could play a role in a host of pathologies. It is possible that SAP is important in coronary heart disease, since SAP has been found specifically localised to human aortic atherosclerotic lesions (Li et al., 1995). That study provided indirect evidence for the presence of amyloid in atherosclerotic plaques, and it has been proposed that SAP exerts pro-atherogenic effects by both accelerating the formation of amyloid fibrils and inhibiting their degradation (Stewart et al., 2007). However, other studies suggest that it is the small oligomeric species present early in the process of fibril formation that are cytotoxic (Novitskaya et al., 2006), so SAP may be carrying out a protective function by converting these into mature amyloid fibres. Further research into SAP is clearly needed to fully understand the role this protein may play in pathogenesis or protection against inflammatory disease.

1.6 ZiCo: a designed switch peptide

This section will introduce the reader to ZiCo, a peptide that was designed to switch between two folded forms upon binding zinc. I will discuss the theory behind the design and the interplay between protein sequence and structure from a design perspective. I will critically discuss previous work carried out on this system and explain how obtaining high-resolution structures is crucial to validating the success of the design.

1.6.1 Protein folding and design

The key to carrying out the multitude of tasks that proteins perform *in vivo* lies in their ability to adopt unique three-dimensional shapes. Proteins are synthesised as linear chains of amino acids, which must fold into a specific conformation to allow their biological function to be carried out. The question of precisely how proteins assume their native structure under physiological conditions is known as the protein folding problem, and it is widely recognised as one of the greatest current challenges in molecular biology.

The search for an answer to the protein folding problem has taken on a new sense of urgency in the post-genomic era because of the drive to exploit the vast amount of information gained from genome sequencing studies. Several initiatives have been launched to complement the human genome project with a map of the human proteome, including the Human Proteome Project sponsored by the Human Proteome Organisation. Indeed, developments in sequencing technology are ushering in the advent of personal genomics, with increasingly rapid turnaround times (days as opposed to years) and the cost of sequencing a single genome predicted to fall below £1000 in the coming years. Structural information is key in order to take advantage of the vast sets of protein identities generated by such projects. Obtaining experimental structures of a

protein is a lengthy and costly process, and may not even be practicable for many proteins due to difficulties in obtaining adequate amounts of material or the amenability of the protein itself to the techniques in question. Furthermore, the true power of such resources lies not in our ability to document the proteins present in life, but in our ability to manipulate them. For this reason, the ability to not only predict protein structures from sequence data alone but also to modify them in a directed way is crucial for progress in biology and medicine.

The protein folding field has been reviewed extensively (Dill et al., 2008; Basharov, 2003; Yon, 2002; Radford, 2000). An understanding of protein folding is of fundamental importance to science since it relates protein function to the genetic code, thus providing the missing link between genotype and phenotype. A principal goal of protein folding studies is to allow any structure to be predicted from its sequence alone. Being able to do this would revolutionise research in the biological sciences and have profound clinical applications. A related goal involves the design and redesign of protein sequences to adopt specific structures or perform specific functions. While there are numerous examples of both *de novo* designed structures and proteins redesigned to possess increased stability or altered binding characteristics, the successful design of novel folds from first principles is still a challenging and complex process.

Protein folding is both spontaneous and reversible, and pioneering studies on ribonuclease by Christian Anfinsen demonstrated that the information required to assume a tertiary fold is contained within the primary sequence (Anfinsen and Haber, 1961). Subsequent work led Anfinsen to propose his thermodynamic hypothesis: that the native state of a protein is the thermodynamically most favoured conformation (Anfinsen, 1973). In 1968, Cyrus Levinthal proposed that proteins must fold via specific pathways (Levinthal, 1968), having shown that it would take a small protein longer than the age of the universe to fold through a random search of conformational space. The fastest proteins are known to fold on microsecond time scales (Eaton et al., 2000), and this large discrepancy between the calculated and observed times for protein folding has become known as the Levinthal paradox. These early observations have

underpinned the field of protein folding, and each addresses a broad area of the protein folding problem: the sequence/structure relationship; protein stability; and folding pathways.

Much work has gone into elucidating the mechanisms by which proteins fold and important insights have been gained from experimental and theoretical approaches. Technological advances have led to the development of new techniques enabling us to probe folding in unprecedented detail, the most notable in recent years being force spectroscopy, which allows the unfolding of single proteins to be measured (Fisher et al., 1999). The emerging ideas have been unified into the now commonly accepted ‘new view’ of protein folding. It has become clear that viewing protein folding in terms of a randomly coiled polypeptide chain that folds into its native, lowest energy state via a single pathway is a gross oversimplification.

Firstly, the notion that unfolded proteins exist as unstructured, random coils has been refuted (Smith et al., 1996). Thus the unfolded state is instead considered to be an ensemble of transient conformations of high free energy. Similarly, the native state does not correspond to a single conformation, but to multiple folded forms that undergo rapid inter-conversion at lower free energy. Furthermore, there is a growing body of evidence to suggest that proteins fold co-translationally, and the validity of the post-translational view of protein folding has been directly questioned (Basharov, 2000). Studies on α -lytic protease have shown that the unfolded and partially unfolded states are lower in free energy than the native, directly contradicting the thermodynamic principle (Sohl et al., 1998), and proteins are also known to fold via multiple pathways (Weissman, 1995).

The folding process is now commonly described in terms of a funnel-shaped energy landscape. The top of the funnel represents the vast number of possible conformations of the unfolded state, the native state is represented as a well at the bottom, and the kinetic and thermodynamic aspects of the folding process are represented as bumps in the surface of the funnel. While this model unites much of our current understanding of the principles behind protein folding, our understanding of

how proteins fold *in vivo* is complicated further by the role of auxiliary proteins such as chaperones that assist the folding process. A discussion of this topic is beyond the scope of this text, but the reader is directed towards recent reviews (Hartl and Hayer-Hartl, 2009; Ellis, 2006).

The study of protein folding has been revolutionised by our ability to engineer proteins and even design them from first principles. This has been made possible both by advances in our understanding of protein folding and by the development of new techniques and tools with which to design and manipulate proteins (for a discussion of recent advances in protein design methodology see Lippow and Tidor, 2007). Protein design allows us to tease out information about the structural stability and functional role of individual parts of the protein structure under scrutiny. In combination with information gleaned from bioinformatics approaches, this can enable us to establish new rules for protein folding and deepen our understanding of the relationship between sequence and structure. Protein design often focuses on producing a single stable structure but it is structural fluidity that underpins the function of many proteins. So it is important not only to understand how to design stable protein folds, but also how to engineer any required flexibility into these structures that will allow novel functions to be created. Indeed, a blossoming area of protein design focuses on proteins whose primary sequence encodes multiple structures, and this will now be discussed.

1.6.2 Switch peptides

Proteins have long been known to be dynamic molecules. The constant thermal motion of atoms can result in small movements such as the flipping of side chains — which may be important for carrying out the biological function — as well as large conformational changes. These movements, in response to ligand binding or changes in the local environment, underpin many normal physiological processes. The implication of conformational changes in various diseases (Carrell and Gooptu, 1998), collectively

known as protein conformational disorders, has led to a renewed interest in the structural flexibility of proteins. The defining event in such diseases is a change in the structure of a protein with no concomitant change in the primary sequence, which results in pathogenicity either through the loss of a biological function or through the acquisition of a toxic activity.

There has been a considerable paradigm shift away from the ‘one sequence, one structure, one function’ hypothesis that was prevalent in the past, because it has since become apparent that a single sequence is able to form more than one stable fold. Examples of sequences that exhibit structural duality abound in naturally occurring proteins (Damaschun et al., 1999; Mezei, 1998; Zhao et al., 2001). A particularly striking example is the bacterial elongation factor EF-Tu, which contains a six-residue sequence that completely converts from an alpha helix when GTP is bound, to beta structure in the GDP-bound form (Abel et al., 1996). There are also numerous examples of designed sequences capable of switching between stable conformations in response to the local polypeptide environment (Minor and Kim, 1996), temperature (Ciani et al., 2002), redox state (Pandya et al., 2004), phosphorylation (Signarvic and DeGrado, 2003), solvent conditions and light (Cerpa et al., 1996).

Of course, ligand binding also induces structural changes in proteins, and there has been a growing trend towards the design of metal-binding peptides that produce a measurable conformational change (Pearce et al., 1998; Shults et al., 2003). Zinc-binding peptides have been a particular focus of recent research, largely because the zinc finger domains on which such designs are based have been extensively studied (Berg and Godwin, 1997). Besides furthering our understanding of protein folding, such designed peptides have considerable practical applications as sensors in quantitative and qualitative analyses of metal ions in environmental and clinical samples, since protein based sensors can offer unrivalled selectivity. Pursuing the design of metal binding protein switches is attractive, then, in terms of both intellectual merit and technology transfer.

Another desirable property of biosensors based on conformational changes in proteins is their intrinsic potential for reversibility. With this in mind, Cerasoli et al., 2005, designed ZiCo, a peptide that switches reversibly between two very different folded forms upon metal binding. The principle behind the design involves the combination of two different sequence motifs into a single sequence. If the design is successful, ZiCo will switch between a trimeric α -helical structure and a folded monomer upon binding divalent zinc, as illustrated in Figure 1.7. The details of the design and its applications will be outlined in the following sections.

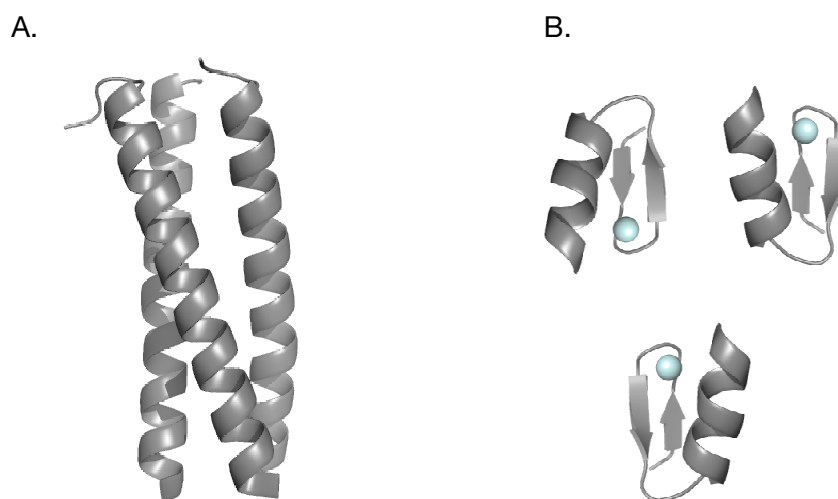


Figure 1.7: ZiCo design principles. A. This trimeric coiled coil structure illustrates the zinc-free state of the ZiCo design. B. A zinc finger structure illustrates the zinc-bound state. The peptide is designed to switch between the two folds when zinc is added to a buffered saline solution containing the peptide.

1.6.3 Design details

The design and biophysical characteristics of ZiCo will be briefly described here, but for a fuller discussion see Cerasoli et al., 2005. The design process involved merging the consensus sequence of a classical Cys₂-His₂ zinc finger and a non-canonical coiled-coil heptad repeat, with a few added adaptations. A classical zinc finger motif contains a single zinc ion pinched between an alpha helix and a beta turn, co-ordinated by two cysteine and two histidine residues (see Figure 1.8). By contrast a trimeric coiled coil consists of three alpha helices coiled around each other like the strands of a rope.

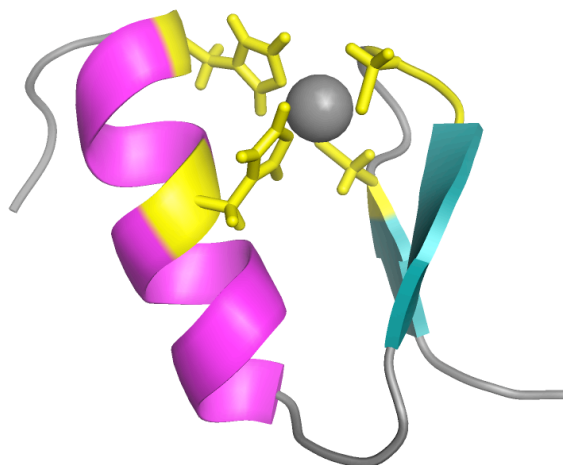


Figure 1.8: Classical Cys₂-His₂ zinc finger. The zinc ion is co-ordinated by two cysteines in the β -turn and two histidines at the C-terminal end of the α -helix. Here the β -structure is shown in cyan, the helix is coloured magenta and the co-ordinating residues are highlighted in yellow.

Coiled coils are characterised by a repeated sequence of hydrophobic and polar residues, known as the heptad repeat. When such a sequence folds as an alpha helix, the hydrophobic residues are located in a seam that runs along one side of the helix. In a coiled coil, the hydrophobic residues (which are distributed along the long axis of the structure) interact in a special way known as ‘knobs-into-holes packing’, whereby a

hydrophobic residue on one helix is buried in a hydrophobic pocket created by residues on another helix. The concept of the heptad repeat and knobs-into-holes packing is illustrated using the helical wheel representation of a trimeric coiled coil in Figure 1.9.

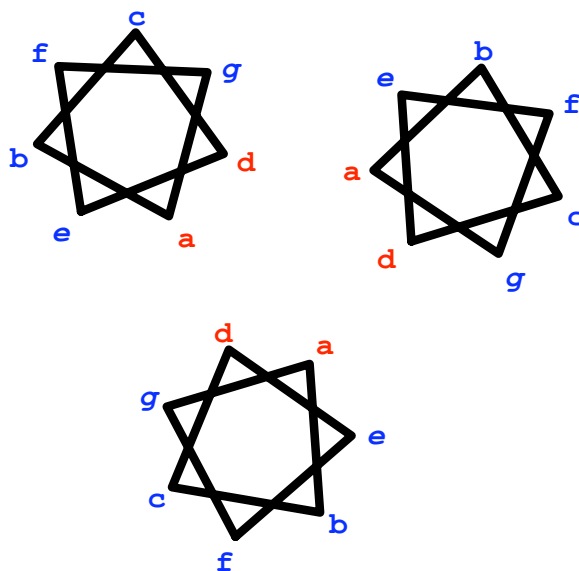


Figure 1.9: The heptad repeat and packing of a trimeric coiled coil model. The seven residues of the heptad sequence repeat are labelled a–g for each of the three helices, with the coiled coil axis running perpendicular to the page. Hydrophobic residues involved in core formation (a and d) are shown in red, while those typically involved in electrostatic interactions (e and g) are shown in italics.

The various features of the ZiCo sequence and their relationship to the structural principles behind the design are depicted in Figure 1.10. Several modifications were made to the raw sequence in order to promote structural flexibility and facilitate the switch between the two folds. Firstly, the cysteine residues of the classical zinc finger have been replaced by histidines, to prevent problems with oxidation. This approach to designing a zinc finger protein has proven to be successful by Hori et al., 2000, who showed that an H₄ zinc finger peptide had a similar structure to the wild type.

Another modification to the sequence is the insertion of a stutter in the heptad repeat, in the region corresponding to the loop linking the alpha helix and the beta

hairpin of the zinc finger. Such stutters have been shown to disrupt the core packing of the coiled coil structures, leading to a local unwinding of the superhelical structure. The stutter was included with the aim of introducing a local region of destabilised core packing to facilitate the switch to the monomeric zinc-bound form.

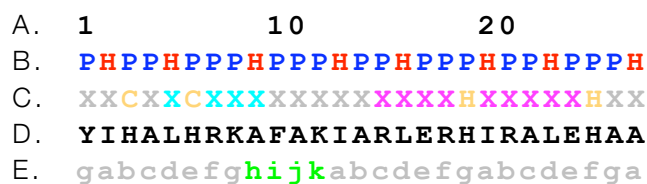


Figure 1.10: Features of the ZiCo sequence showing the design principles. A. Residue number. B. Pattern of hydrophobic (red) and polar residues (blue). C. The classical zinc finger consensus sequence with conserved residues highlighted in yellow, showing regions of α -helix (magenta) and β -strand (cyan). D. ZiCo sequence. E. The heptad repeat pattern showing the position of the stutter from residues 9–12.

In solution, the peptide undergoes a conformational change on binding divalent zinc that can be monitored by circular dichroism (CD) spectroscopy and a change in oligomerisation state that can be followed by analytical ultracentrifugation. Interestingly, a very similar system to ZiCo has been described more recently (Ambroggio and Kuhlman, 2006). The computationally designed peptide Sw2 can be switched from a Cys₂-His₂ zinc finger-like fold to a trimeric coiled-coil fold, in response to pH or the presence of transition metals. This system could provide a valuable model for comparison with ZiCo. The ZiCo peptide as described here was also used as a template for designing a family of similar peptides based on the same core principles. While the published data on the original ZiCo are indicative of a switch between two alternate stable folds, high-resolution structures are the only way of confirming the success of the design of this and the other ZiCo peptides.

1.7 Project background

1.7.1 C1: purification, characterisation and crystallisation

Previous studies on complement component C1 have generally involved purifying the complex by exploiting its affinity for immunoglobulins or other ligands (Medicus and Chapuis, 1980; Schifferli and Steiger, 1985). Often, the individual C1 subcomponents are separated during the purification procedure by the addition of EDTA, and the complex is reconstituted from its purified constituents (Strang et al., 1982; Tseng et al., 1997). However, it is also possible to gently precipitate the entire assembly from human serum using polyethylene glycol, to give a functionally pure extract (Neoh et al., 1984). Given that the entire structure of the complex has not yet been solved, using a novel purification procedure introduces another variable into the search for a crystal structure of native human C1.

Of further interest is the specific role that calcium — and possibly other divalent metals — play in the structure and function of C1 under various conditions. Understanding this relationship has implications both for the search for crystal structures of the native complex and for revealing the full picture of its biological function. An investigation into the effect of various metal ions on the assembly and activity of the C1 complex was carried out some decades ago (Ziccardi, 1983), but the study was not exhaustive. Iron, copper and selenium are all present as trace elements in human blood and increases in copper concentrations are generally accepted to be associated with sites of infection and inflammation. Examining the effects of physiologically relevant conditions on native human C1 may help to further our understanding of this key immune complex, and a model of the entire assembly would contribute greatly to the wealth of structural information already available.

1.7.2 SAP: investigating metal and ligand binding

Serum amyloid P component can readily be purified and crystallised using established methods. Several crystal structures of SAP in complex with various ligands have been solved, taking us some way to understanding the interactions of this protein in the context of immunity. However, there has been little investigation into the effect of metal ions other than calcium on the structure of SAP at high resolution.

In the study by Thompson et al., 2002, the calcium ion in the second calcium binding site could be replaced by cerium, suggesting that under certain conditions other ions might bind to SAP. During inflammation various physiological changes occur, including an increase in acidity and the concentration of copper ions in the local cellular environment. Several studies have demonstrated a link between elevated serum copper levels and cardiovascular disease (Malek et al., 2006; Leone et al., 2006). Increased serum copper levels have also been shown to be characteristic of Alzheimer's disease (Squitti et al., 2005). This supports the theory that these diseases are caused by an initially beneficial inflammatory response that then develops into a chronic pathological process. One possible way in which this may occur is via positive feedback loops that reinforce inflammatory signals in the absence of an appropriate stimulus. It is possible that metal ions could contribute to this process by modulating the function of certain proteins associated with regulating the inflammatory response.

With respect to SAP, there are interesting observations that hint at just such a process. Under conditions such as those present in inflamed tissue, SAP binds to the bacterial cell wall polysaccharide preparation zymosan much more strongly than under normal physiological conditions (Potempa et al., 1985). It is also possible that copper ions interact with SAP in a way that affects the affinity of the molecule for its various ligands. Both SAP and C1 depend on calcium ions for structural stability and function. An investigation into the calcium and copper binding properties of these two proteins under relevant conditions could have interesting implications for our understanding of their structure and biological function in health and disease.

1.7.3 ZiCo: design validation

The ZiCo peptides are *de novo* designed proteins, and the search for high-resolution structures to verify the success of their design was a collaborative project carried out in conjunction with the laboratory of Professor Dek Woolfson (University of Bristol, UK). Prior to this work, ZiCo had been extensively characterised in solution studies, the results of which are summarised in Table 1 (see p. 40). In addition, the original ZiCo peptide had been crystallised — and several data sets had been collected — but attempts at solving the structure by molecular replacement had not yet proved successful.

Owing to the nature of the system, the solution behaviour of ZiCo is inherently less straightforward than that of a typical globular protein. Consequently, the crystallisation landscape of ZiCo in its free and zinc-bound forms is also much more complex. As can be seen in Table 1, the melting temperature of the trimer increases with the concentration of peptide in solution. This suggests that in the absence of zinc, not all the peptide is present in its trimeric state but that there exists an equilibrium between trimers and free peptide in solution. As more peptide is added, this shifts the position of equilibrium in favour of the trimeric form.

In addition, precipitate was observed at high peptide concentrations in the presence of excess zinc and phosphate buffers. It was speculated that this may be due to the formation of zinc hydroxide at around neutral pH, while in other buffers such as Tris, the excess zinc may be sequestered by buffer molecules. Thus both the peptide concentration and the concentration of other solutes can have important consequences for the structure of ZiCo in solution. This must be considered when designing crystallisation experiments with a view to obtaining high-resolution structures of the peptide in order to lend force to the rationale behind its design.

	Zinc	
	<i>None</i>	<i>Equimolar</i>
Oligomerisation state	Trimer	Monomer
% helix	50	75
K_d	<i>Trimer association:</i>	<i>Zinc binding:</i>
	10–100 μ M	3–5 μ M
T_m	<i>Concentration dependent:</i>	
	50 μ M	10 °C
	100 μ M	14 °C
	200 μ M	18 °C
		23 °C

Table 1: Biophysical characteristics of Zico1. Data taken from Cerasoli et al., 2005. The table shows various features of the ZiCo1 peptide in the presence and absence of equimolar zinc concentrations. In the absence of zinc, the peptide forms trimers with a K_d of 10–100 μ M and a melting temperature that is concentration dependent. With equimolar zinc, the peptide forms monomers with a K_d of 3–5 μ M for zinc binding and a T_m of 23 °C.

1.8 Project aims

1.8.1 C1: purification, characterisation and crystallisation

The aim of my work on complement component C1 is to develop a purification scheme to isolate the native protein complex from human plasma using a combination of PEG precipitation and chromatographic techniques. The purification strategy will be optimised for structural studies by X-ray crystallography, with an emphasis on achieving high purity and homogeneity of the purified product. Once purified, the native C1 will be analysed under physiologically relevant conditions to test the effect of various metal ions and other relevant factors on the structure of the complex. Finally, the crystallisability of complement component C1 will be explored in the hope of obtaining usable crystals.

1.8.2 SAP: ligand binding and co-crystallisation

With regards to SAP, the aim of my work is to investigate the ligand binding properties of this molecule. SAP will be purified from human plasma using established methods and the quality of the purified product will be analysed for its suitability for use in structural studies. The metal-binding properties of SAP will be examined using the ability of various non-metal ligands to disrupt the calcium-dependent aggregation of SAP. This work aims to lay the foundations for other projects, including the interaction and co-crystallisation of SAP with metal ions other than calcium.

1.8.3 ZiCo: design validation

Finally, the goal of my work on ZiCo is to provide high-resolution structural information about the system in order to validate the design. The original ZiCo peptide has previously been crystallised, but no structure has yet been found. The aim is to conduct an investigation into the crystallisability of two ZiCo peptides and build upon the available data in the pursuit of a high-resolution structural model of both the free and zinc-bound forms.

Chapter 2

Background to methods

2.1 Overview

The following chapter outlines the general methods employed and discussed in this thesis. Since the main focus of this work is the use of X-ray crystallography to investigate the proteins under study, this chapter will begin with a general introduction to the technique, before examining the different steps in the progression towards solving a structure in detail. One of the crucial requirements for structural studies is a highly purified protein sample, so part of this chapter will deal with the process of obtaining a suitable starting material for structural investigations by chromatographic methods. The remainder of the chapter will explore a variety of other techniques that provide a wealth of complementary information on the purity of a protein sample and its structural characteristics.

There are several techniques available to probe the structure of biological macromolecules, but X-ray crystallography is generally regarded as the gold standard in terms of the accuracy of the structures obtained and the range of situations in which it can be applied. While there have been interesting developments in the field relating to the possibility of imaging single protein molecules at atomic resolution by X-ray scattering (Shintake, 2008), the current state of technology requires the use of crystals to generate a measurable diffraction signal.

Obtaining a high-resolution protein structure by the crystallographic method essentially involves interpreting the pattern of X-rays diffracted by the array of identical molecules that occupy a crystalline lattice, in order to produce a map of the electron density within the structure. In the general set-up, a crystal is rotated within an incident beam of monochromatic X-ray radiation and the diffracted rays are recorded as a series of two-dimensional images. Using a computer to apply the mathematical method of Fourier transforms, the intensities and positions of the reflections are combined with information about their phases and converted into a three-dimensional electron density map. A model can then be generated by making use of prior

information about the specific protein under study (and the structure of proteins in general) and progressively fitted to the experimental data until a satisfactory model of the structure is achieved.

Compared with the task of determining the structure of a small molecule compound, for which X-ray crystallography is also an established technique, the process of solving a protein structure is complicated by several factors. Firstly, obtaining the protein product in a suitably pure and stable form (and in sufficient quantities) for use in crystallisation trials can be problematic. Secondly — in contrast to the simple evaporation procedures typically used by chemists — the production of protein crystals suitable for diffraction studies is generally a very intensive task that can consume vast amounts of material and considerable lengths of time. And finally, the complexity and sheer size of the molecules in question mean that the computational techniques for determining the structure from the diffraction data are also more convoluted.

Each stage in the procedure, from preparing a pure protein sample to refining the structural model, has its own peculiar difficulties and must be tailored to the protein in question. Particularly in the case of entirely novel structures, it is not possible to predict from the outset exactly what strategies to employ at each stage, which means that in many cases structure solution by X-ray crystallography remains a rather empirical procedure. The following sections will describe the various steps in the progression towards solving a protein structure in detail.

2.2 Protein purification

One of the most crucial steps in preparing a protein for crystallisation purposes is purifying the protein itself. In contrast to many standard biochemical applications, structural studies generally require a highly pure and conformationally homogeneous sample. As a result, a multi-step procedure is generally employed to achieve the level of purity required for crystallisation. There are a variety of methods available to purify proteins from different sources and the major techniques used in this work will be outlined in the following section. In the process of designing a purification scheme, the nature of the protein in question must be considered, for example, any ligand binding properties may be exploited in an affinity step. Furthermore, the final purified product is best finally reconstituted and stored in a manner that maintains the protein in a structurally stable state.

2.2.1 Precipitation

When purifying a protein from a complex source, such as cell lysates or bodily fluids, a common first step is to precipitate a fraction of the soluble proteins that contains the protein of interest. This can be accomplished by adding neutral salts, such as ammonium sulphate, or polymers such as polyethylene glycols (PEGs) to an appropriate concentration that results in precipitation of the target protein from solution. Proteins precipitate when the repulsive electrostatic forces between them are reduced or when hydrophobic patches become exposed on the protein surface, causing aggregation and precipitation.

Adding salts to a protein solution results in the bulk water becoming associated with the salt ions, meaning that less water is available to partake in the solvation layer surrounding the proteins. Similarly, polymers attract water molecules away from the solvation layer around the protein, increasing protein-protein interactions and enhancing precipitation. Compared with other methods of precipitating proteins for further purification, the advantage of using salts and polymers is that they are less likely to cause irreversible denaturation. The precipitated fraction will be enriched in the protein of interest but will generally still contain a number of other proteins, which must be removed by further purification steps, for example by separation using chromatography.

2.2.2 Chromatography

Chromatographic techniques allow the separation of protein mixtures by passing a sample dissolved in a mobile phase, such as an aqueous medium, through a stationary phase of some sort, separating the target protein from other molecules in the mixture based on differential partitioning between the mobile and stationary phases. Chromatography can be preparative or analytical. The former uses large amounts of sample with the aim of purifying the desired components from the mixture, while the latter usually uses smaller amounts of material and can be used to analyse the proteins in a sample in terms of various properties of the components and experimental conditions.

2.2.2.1 *Affinity chromatography*

Affinity chromatography separates proteins on the basis of a reversible, specific interaction between the target protein and a ligand coupled to the chromatography matrix. It is often used alone or towards the beginning of a multi-step purification procedure, since it offers high selectivity and often also high capacity for the target protein, thus allowing considerable quantities to be extracted directly from complex mixtures.

An affinity chromatography procedure involves selecting an appropriate ligand and coupling it to a suitable medium such that it is still free to bind the protein of interest. The sample to be separated is applied to the column and allowed to bind to the immobilised ligand, after which the unbound contaminants are washed from the column. The remaining bound sample is eluted from the column either specifically, by introducing a competing ligand, or non-specifically by changing the pH, ionic strength or polarity so that the interactions between the sample and the matrix are weakened.

Finally, the purified sample may be further purified by other methods or exchanged into a suitable buffer for further studies, since the elution buffer may be unsuitable for downstream applications due to the presence of additives. The major factors that affect the outcome of the separation are the choice of ligand and coupling procedure, the binding and elution buffers, and the flow rate during these two major steps. By carefully optimising these parameters, a particular protein can be purified from a complex mixture or information about its binding properties can be elucidated.

2.2.2.2 Ion exchange chromatography

A further chromatographic technique is ion exchange chromatography (IEX), where molecules are separated based on their net surface charge. The overall charge of any given protein will change depending on the pH of the surrounding medium and the charge density and charge distribution varies considerably from one protein to another. IEX exploits the fact that the relationship between pH and net surface charge is particular to each protein. The separation of the various components of a sample occurs as a result of reversible interactions between the charged proteins and the oppositely charged IEX medium, which can be controlled to give preferential binding and elution of specific molecules. At a pH above its isoelectric point, a protein will bind to a positively charged medium or anion exchanger, whereas at a pH below its pI, a protein will bind to a negatively charged medium or cation exchanger.

Depending on the pI of the protein and the pH at which the separation is to be carried out, a suitable medium must therefore be selected. Since the binding of molecules to the column is affected by ionic strength and pH, the samples must generally be prepared in the same buffer as that used during the binding step of the separation procedure. As with affinity methods, the sample is loaded onto the column and any unbound substances are washed away. By changing the ionic strength (or alternatively the pH) of the buffer passing through the column, the bound

substances can be eluted. Changing the pH directly affects the interactions between the proteins and the medium, whereas increasing the ionic strength causes salt ions or other solutes to compete with the bound molecules for charges on the surface of the stationary medium. Proteins with a low net charge will elute from the column first, since they have fewer interactions with the medium, while those with a high net charge will be retained for longer on the column and elute last. By varying the changes in ionic strength using different types of gradient during the elution, the separation can be fine-tuned to give an optimal purification.

2.2.2.3 Size exclusion chromatography

Yet another purification technique useful when working with proteins is size exclusion chromatography (SEC), also known as gel filtration chromatography, in which particles are separated on the basis of differences in size. The sample to be separated is suspended in a suitable buffer and passed through a column packed with a porous and chemically inert medium, typically agarose, dextran or polyacrylamide beads. Smaller molecules are able to penetrate the interior of the beads, while large particles are excluded from the beads entirely and so only have access to a part of the column volume, known as the void volume. The principle behind SEC is that molecules of different sizes will pass through the column at different rates, based on the extent to which they are able to penetrate the stationary phase. Large molecules, having access to a smaller volume, elute from the column first, with smaller proteins eluting according to their hydrodynamic properties.

While SEC is perhaps the simplest way of separating a protein mixture, a number of factors have to be considered. In order for separation by SEC to be successful, the particles to be separated must be loaded onto the column simultaneously. In practice, this means that the sample must be applied in a small volume relative to the volume of the column. In addition, the medium must be selected

to prevent any specific interactions with the sample components and to give good resolution over the range of molecular weights required. The size of the pores in the matrix define an upper and a lower boundary for the size of molecules that can be reliably separated. Above the exclusion limit, large particles elute together with the void volume, and below the permeation limit, where small molecules can completely penetrate the medium, smaller particles elute as a single peak.

A column can be calibrated against proteins of known molecular weights and thus the molecular weight of the sample protein can be estimated. Strictly, the elution of proteins from a SEC column is governed by the hydrodynamic properties of the particles, which reflect both the size and shape of the molecule. Accordingly, a globular protein and an extended polymer of the same molecular weight will behave differently during the separation. Compact particles have a greater probability of entering the pores of the matrix, since extended particles must be oriented correctly in order to gain access to the hidden volume. Thus SEC can be used to distinguish between folded and unfolded forms of the same protein, since the folded form is likely to elute much later. Aside from its application for studying protein structure, SEC is generally employed in the final stages of a purification scheme, as a polishing step. It is particularly convenient as it allows the protein to be exchanged into a buffer suitable for storage or further work, thanks to the size difference between protein and low-molecular-weight solute molecules.

2.2.3 Protein concentration

Once a protein has been purified, it is often necessary to concentrate the sample for further studies, particularly if it is to be used in crystallisation trials. This was accomplished here by the use of centrifugal filters, which contain a microporous membrane that retains molecules above a certain molecular weight while allowing solvent and smaller particles to pass through. Like SEC, ultrafiltration also allows buffers to be exchanged quickly and easily when compared to other methods such as dialysis.

2.3 Protein characterisation

Proteins can be characterised by various methods, and those relevant to the investigations presented in this thesis will be described in the following sections. Often the aim of a biochemical study is to assess the function of a protein. This does not require the preparation of very pure extracts. As long as inhibitors and other destructive or damaging contaminants are removed or their effects minimised, functional assays can be performed using simple homogenates without the need to remove other proteins present in the sample. However, for the purposes of obtaining high-resolution protein structures by X-ray crystallography or other means, it is necessary to ensure as pure a preparation as possible. In addition to confirming the success of a protein purification procedure, the techniques employed in this work provide complementary information about the solution structure and behaviour of the systems under study.

2.3.1 Electrophoresis

Electrophoresis is an invaluable tool in protein biochemistry, both for the purposes of purification and as a way of analysing biomolecules quickly and cheaply. Charged particles — proteins, for instance — will move through a porous medium when an electric field is applied. The relative rates of migration depend on a number of factors, including net charge, molecular mass and the hydrodynamic radius of the molecules, as well as the properties of the medium through which the proteins are migrating. This is the principle behind all electrophoretic methods and the present section will outline specific techniques employed in this work.

2.3.1.1 *Native PAGE*

Perhaps the simplest way to analyse proteins by electrophoresis is the method known as native polyacrylamide gel electrophoresis (PAGE). As the name suggests, the electrophoretic separation of a protein mixture is performed under non-denaturing conditions, allowing the native state as well as protein-protein interactions to be maintained. There are two main flavours of the technique: blue native PAGE, which uses the Coomassie blue dye to impart a net negative charge to the proteins; and clear native PAGE, where the intrinsic negative charge of most proteins provides the necessary electrostatic attraction to the anode and hence the end of the gel. Of course, if a horizontal system is used and the sample loading wells are placed in the centre of the gel, positively charged proteins will also migrate through the gel towards the cathode, whereas in a vertical system these would be simply lost to the upper buffer chamber. The net charge of molecules in solution can be manipulated by varying the pH.

Only clear native PAGE was used in this work, as Coomassie binding to proteins can cause the dissociation of protein complexes. However, because of the large

size of the proteins being analysed, an agarose gel — with a much larger pore-size — was substituted for the polyacrylamide gel. In a typical procedure, an agarose gel was cast by dissolving the required amount of agarose in 100mM Tris, pH8.5 and using the same buffer as a running buffer. Samples were also loaded in the same buffer containing 20% glycerol and 0.010% bromophenol blue, to monitor the progress of the run. Since the majority of proteins are acidic, i.e. having a $pI < 7$, the high pH of the running buffer ensures a slight negative charge. Of course, the buffer system can be adjusted to accommodate the specific properties of the protein of interest or to achieve specific desired effects.

2.3.1.2 SDS-PAGE

Separation of proteins on the basis of their molecular masses alone is achieved quickly and easily by SDS-PAGE, which is one of the standard repertoire of biochemical techniques available for studying proteins. Briefly, the protein sample is heated to 100°C in the presence of SDS, a detergent which binds to the polypeptide backbone in a constant ratio and coats the protein in an overall negative charge, effectively masking the intrinsic charge of the protein. Thus the proteins within the sample become linear structures with a uniform charge density and the electrophoretic mobilities of these proteins will be a linear function of the logarithms of their molecular weights.

The denatured proteins can then be separated on a polyacrylamide gel immersed in buffer. As the negatively charged molecules migrate in the electric field through the pores of the gel, smaller molecules are more mobile and thus travel faster than larger molecules. The progress of an SDS-PAGE run can be monitored using a dye, bromophenol blue, which travels through the gel matrix ahead of the smallest proteins because of its small size. Terminating the run once the dye front reaches the bottom of the gel ensures that none of the proteins are lost into the surrounding buffer.

The most common buffer system used for SDS-PAGE is the discontinuous Laemmli method. Discontinuous buffer systems, which use a different buffer for the gel and the tank, concentrate the proteins into a very tight band early in the run, resulting in increased resolution of the resulting separation. However, continuous systems, where the same buffer is used both in the gel and the tank, may also be used, although this is much less common. The results of an SDS-PAGE separation can be visualised using one of the staining methods outlined below.

2.3.1.3 *Staining*

Since most proteins are invisible during the electrophoresis run (exceptions being proteins that are naturally coloured or have been labelled with some dye) the progress is monitored using the small molecule dye bromophenol blue, which migrates ahead of any proteins. However, after the run has been completed, the proteins themselves must be visualised by some means. There are various methods to determine the positions of proteins separated by electrophoresis within the gel, the most common being staining with Coomassie Brilliant Blue.

The Coomassie family of dyes are small-molecule dyes that bind non-specifically to most proteins via interaction with certain residues (primarily lysine, arginine and histidine), as well as through Van der Waals and hydrophobic interactions. The amount of dye bound to each molecule is roughly proportional to the number of positive charges on the protein. Free amino acids, peptides and proteins with a molecular weight less than around 3kDa do not stain with Coomassie reagents. However, for most proteins, Coomassie staining offers a quick, simple and cost-effective method of staining. In addition, it is compatible with mass spectrometry and can be performed prior to further staining with the more sensitive silver staining method.

Two Coomassie staining protocols were used in this work, which will briefly be explained here. The standard protocol involves fixation and staining of the protein

bands in a solution of Coomassie Blue R-250 in methanol and acetic acid for several hours to overnight. The staining procedure can be performed rapidly on polyacrylamide gels by heating the gel and staining solution using a microwave. Since agarose may melt at temperatures above $\sim 60^{\circ}\text{C}$, this rapid staining method cannot be used with agarose gels. Stained gels are subsequently incubated in de-staining solution, which removes unbound dye from the gel so that the stained protein bands can be seen.

The second method differs from the first in that instead of using dissolved dye, the staining solution is a colloidal suspension of Coomassie Blue G-250. Unlike the standard method, the colloidal Coomassie does not appreciably stain the gel itself but only the protein bands, and therefore does not require a de-staining step. Protein bands appear rapidly (< 1 hour) and the method has been reported to detect as little as 1 ng of protein (Candiano et al., 2004), compared with a detection limit of around 100 ng for the standard protocol.

Silver staining of proteins provides a comparable sensitivity to the colloidal Coomassie stain outlined above, but at considerably greater expense and effort. The principle behind silver staining is that proteins bind silver ions, which can be reduced to metallic silver to give a visible image, much like photographic film. Proteins must first be fixed in the gel and any interfering substances must be removed by washing before the gel is sensitised to promote silver reduction. The gel is then suffused with silver solution and finally the image is developed. There are two general methods of silver staining: one uses a silver nitrate solution and the other makes use of a silver-ammonia complex. However, many variations on these basic protocols exist. The choice of which silver staining protocol to use depends on various factors, such as the formulation of the gel being stained and any requirement for downstream applications of the stained spots. For example, fixation with formaldehyde or glutaraldehyde dramatically improves the quality of staining but is incompatible with mass spectrometry (Chevallet et al., 2006).

2.3.2 Mass spectrometry

Mass spectrometry is a technique that can be used to analyse elemental composition, obtain accurate molecular masses or gain information about the structures of biomolecules based on measurements of their mass to charge ratio (m/z). Mass spectrometry works by first vaporising and ionising the sample molecules to obtain gaseous ions. There is a plethora of ways in which this can be achieved, the two most commonly used for biological samples being electrospray ionisation and matrix-assisted laser desorption/ionisation. Once the gaseous ions have been created, they are accelerated by an electric field and pass into a mass analyser, which uses electromagnetic fields to separate the ions based on their m/z ratio. Again, there are numerous types of mass analyser, ranging from simple sector and time-of-flight instruments to more complex ion trap and cyclotron resonance devices. Finally, the induced electrical charge or current is measured when the ions interact with a detector, and the signal intensity — which is closely correlated with the relative abundance of each ion — is plotted against m/z to give the mass spectrum of the sample.

Mass spectrometry is quantitative, accurate, and capable of extremely high resolution, able to discriminate between molecules that differ by a single atomic mass unit. This makes it a powerful tool with which to probe proteins and their interactions. The applications of mass spectrometry to the study of biological macromolecules are numerous and varied; this is presumably a reflection of the diversity of the available instrumentation. Mass spectrometers are modular instruments, with different sources and detection methods available depending on the sample and the type of experiment being performed. One particularly useful feature from the point of view of a structural biologist is the ability of mass spectrometry to be coupled seamlessly to other techniques, such as liquid chromatography.

Another interesting set-up is known as tandem mass spectrometry (MS/MS), where several mass analysers are set up in series. Typically, sample molecules are separated and analysed in the first mass analyser, and a subset is selected and directed

towards a second mass analyser, filled with an inert gas. Here the sample ions collide with gas atoms and are fragmented to produce smaller ions, which are analysed in the final mass analyser. Alternatively, tandem mass spectrometry can be performed over time in a quadrupole ion trap, without the need for further mass analysers.

Perhaps the most routine use of mass spectrometry in the biosciences is to confirm the identity of a protein. Tandem mass spectrometry can be used to identify proteins or protein fragments, but its use is not as widespread as other methods of protein identification, such as peptide mass fingerprinting (see below). During the fragmentation process, proteins mainly break at peptide bonds, essentially giving rise to a number of shorter peptides that differ by only a few amino acids. By comparing the results with a database of predicted masses for many given peptide sequences, the original sequence of the protein can be identified with an associated probability score.

2.3.2.1 MALDI-TOF and peptide mass fingerprinting

The identification of proteins and other substances is routinely achieved with ease and speed by a type of mass spectrometry known as MALDI-TOF. Here, the sample is mixed with an acidic matrix and a small amount is deposited onto an inert surface and dried. A laser is used to vaporise and ionise the sample, and the particles are measured in a time-of-flight detector. This type of mass analyser accelerates the ions through the same potential, which ensures that the kinetic energies of the particles are equal. As a result, the velocity of each particle depends solely on its mass, with lighter particles travelling faster than heavier ones. By recording the time each ion takes to reach the detector, it is possible to obtain very accurate mass measurements. If the molecular mass of the protein is known, this can be used as a means of identification.

Of course, this method will only work with pure proteins whose molecular mass is documented and which are not modified in any way. Unfortunately, this represents a rather small subset of all proteins. However, a method known as peptide mass

fingerprinting allows proteins to be easily identified from complex mixtures. Typically, a sample is separated by two-dimensional gel electrophoresis, in which proteins are separated first by IEF and then by SDS-PAGE in the second dimension. The combination of these two techniques — which each have a high resolving power — usually results in a gel where each protein is distinct from every other protein, even when there are many hundred proteins present in the original sample.

Sections of the gel containing the proteins of interest are excised and digested proteolytically. The resulting peptides are then analysed and their experimentally determined masses are compared against a database that contains the masses predicted from the digestion of a number of known proteins. Such a reference database can be constructed by computationally cleaving a list of known protein sequences, for example from the SWISS-PROT database. If a protein in the reference database gives rise to a significant number of peptides that match the experimental masses, the inference is that this protein was present in the sample.

2.3.2.2 Mass spectrometry of protein complexes

Of course, mass spectrometry in protein science is not limited to simply identifying proteins of interest. Much more interesting information can be obtained using an electrospray source, which allows the study of liquid samples. Here, an aqueous solution is drawn into a fine capillary that has a voltage applied to it. The electric field distorts the shape of the liquid surface into a cone, and at a sufficiently high voltage, a jet of liquid is emitted from the cone tip, resulting in a fine spray of droplets. As they move through the source, the solvent evaporates, leaving behind gaseous ions that can then be introduced into the mass analyser.

In contrast to MALDI-TOF, electrospray ionisation tends to produce multiply-charged ions; and since mass spectrometry measures m/z , instead of mass directly, this brings the analysis of larger proteins well within the mass range of many instruments.

Furthermore, the electrospray process is ideal for studying non-covalent protein complexes, since it is capable of transferring them into the gas phase without disrupting the interactions that stabilise them. This makes it a powerful tool for structural biologists, demonstrating how mass spectrometry can extend far beyond its common use for simple protein identification.

As just noted, electrospray ionisation maintains non-covalent macromolecular complexes intact in a mass spectrometer. This means that research ranging from the study of simple homo-oligomers to heterogeneous assemblies of mega-Dalton proportions can be carried out using the advantages that mass spectrometry offers, such as the need for only limited quantities of sample in experiments and the high accuracy and resolution of the resulting data. Thus mass spectrometry is both a complement to standard structural biology techniques like X-ray crystallography, but also possesses unique attributes that make it a valuable tool in its own right.

Examples of mass spectrometry and its application in studies on the architecture and biogenesis of macromolecular complexes and protein interaction networks can be found in recent reviews (Sharon and Robinson, 2007; Gingras et al., 2007; Heck, 2008). Mass spectrometry also finds specialised applications in identifying and investigating metalloproteins (Shi and Chance, 2008).

2.3.3 Dynamic light scattering

Dynamic light scattering (DLS) is another technique that can be used to determine the size distribution of small particles in solution. It is non-destructive, and — in contrast to the methods outlined so far — the sample can be reclaimed, making it an excellent tool when working with precious protein samples. DLS can be used to monitor aggregation, determine the presence of various oligomerisation states, give an estimate of molecular shape and predict the crystallisability of a protein. Experiments are quick and easy to perform and reactions can be monitored in real time.

2.3.3.1 *Scattering theory*

When light hits small particles in solution, the light scatters in all directions, provided that the particles are small compared with the wavelength of the incident radiation. This elastic scattering — where the kinetic energy of the particles is conserved — is known as Rayleigh scattering. If the light source is monochromatic and coherent, for example a laser, then the scattering intensity can be observed to fluctuate over time. These intensity variations are due to the fact that the small particles are in constant motion, so the distance between the scatterers in solution is constantly changing. The scattered light then undergoes either constructive or destructive interference by the surrounding particles and the resulting variation in intensity contains information about the movement of the scatterers.

Once intensity data have been collected, the dynamics of the particles are derived by cross-correlating the intensity trace recorded during the experiment with itself to generate an autocorrelation function that relates the measured intensities at various time points. Essentially, at short time delays, there is a high correlation because the particles have not moved far from their initial positions. As the time delay

increases, the correlation decays exponentially to zero. This decay is related specifically to the diffusion coefficient of the particles, which in turn is related to their hydrodynamic properties.

Finally, computational methods based on calculations of assumed distributions are used to fit the correlation function. If the sample contains particles that are uniform in size, then the decay is simply a single exponential (cumulants analysis), while a population of differently sized particles can be fitted by more complex procedures to the optimal number of exponentials (multimodal analysis). Thus, various properties of the sample can be derived, including the size, shape and relative distributions of the particles in the solution being analysed.

Scattering theory typically assumes that the particles in question are spherical. However, an elongated molecule will give a higher apparent radius than the corresponding sphere. This has important implications when using DLS to study proteins. Globular proteins may be well approximated by spheres, but extended or linearly aggregated proteins will behave differently in solution and this must be considered or even exploited in the experimental design. Using commercially available software, the axial ratios of spheroid particles and their associated Perrin factors can be estimated, allowing quite complex structural information to be extracted about the solution structure and behaviour of proteins.

2.3.3.2 Interpreting DLS data

The output of a DLS instrument typically includes the Z-average size and polydispersity index (PDI) of a sample, as well as various particle size distributions. The Z-average is the mean average size determined from a cumulants analysis (decay fitted to a single exponential), while the PDI is a measure of how well the data fit this model. In practical terms, the PDI is the relative variance of a hypothetical Gaussian distribution centred on the Z-average size. A low PDI indicates that the sample

contains a single population of particles with a narrow distribution of sizes, while a high PDI indicates either a wide distribution (assuming a single population) or that there are multiple populations of differently sized particles in the sample.

For samples with a low PDI (or that are known to be a single population) the cumulants analysis is preferable, since it makes no assumptions about the noise in the data (see Figure 2.1). However, for particles with a high PDI the results derived from a multimodal analysis are generally more meaningful, since there are likely to be several distinct populations of differently sized particles within the sample (see Figure 2.2). The choice of which analysis to report is best made by using all of the available information about the sample under study.

The scattering intensity of a particle is proportional to the square of its molecular weight, in other words larger particles scatter much more strongly than smaller particles. Therefore, even a small number of very large particles (dust, aggregates) can significantly skew the results of the cumulants analysis. The primary size distribution obtained from a DLS experiment is based on scattering intensity, but this can be transformed into a distribution based on the volume of the particles (assuming the particles are spherical). This can then be further converted into a distribution based on particle number. Intensity distributions are typically the most reliable, since any errors are compounded by the assumptions made in the process of converting to volume and number distributions. An example of the various distributions is given in Figure 2.3.

	Diam. (nm)	% Intensity	Width (nm)
Z-Average (d.nm): 63.65	Peak 1: 65.22	100.0	12.72
Pdl: 0.029	Peak 2: 0.000	0.0	0.000

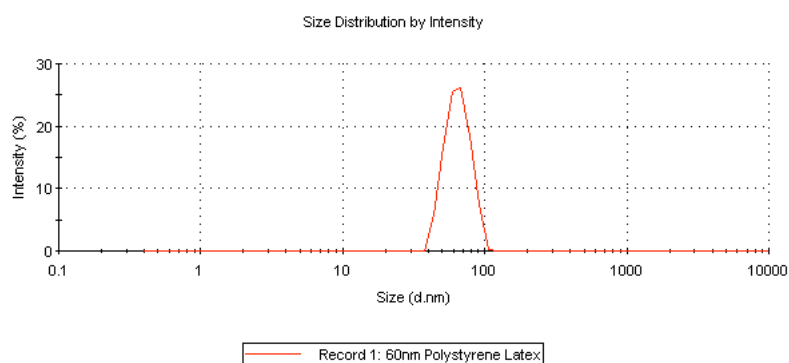


Figure 2.1: Size distribution of 60nm latex beads. The cumulants analysis (red) gives a Z-average size of 63.65nm, with a low PDI of 0.029. These results are in agreement with the multimodal analysis, which shows that the distribution is monomodal with a single peak at 65.22nm.

	Diam. (nm)	% Intensity	Width (nm)
Z-Average (d.nm): 173.7	Peak 1: 228.1	88.5	57.37
Pdl: 0.196	Peak 2: 64.11	11.5	8.741

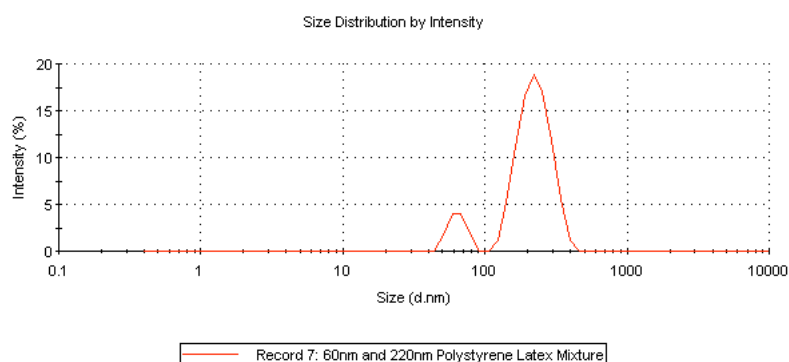


Figure 2.2: Size distribution of a mixture of 60nm and 220nm latex beads. The Z-average size is 173.7nm, with a higher PDI of 0.196. Multimodal analysis shows a smaller peak at 64.11nm and a larger peak 228.1nm in the size distribution by intensity. In this case, the results of the multimodal analysis more accurately describe the sample.

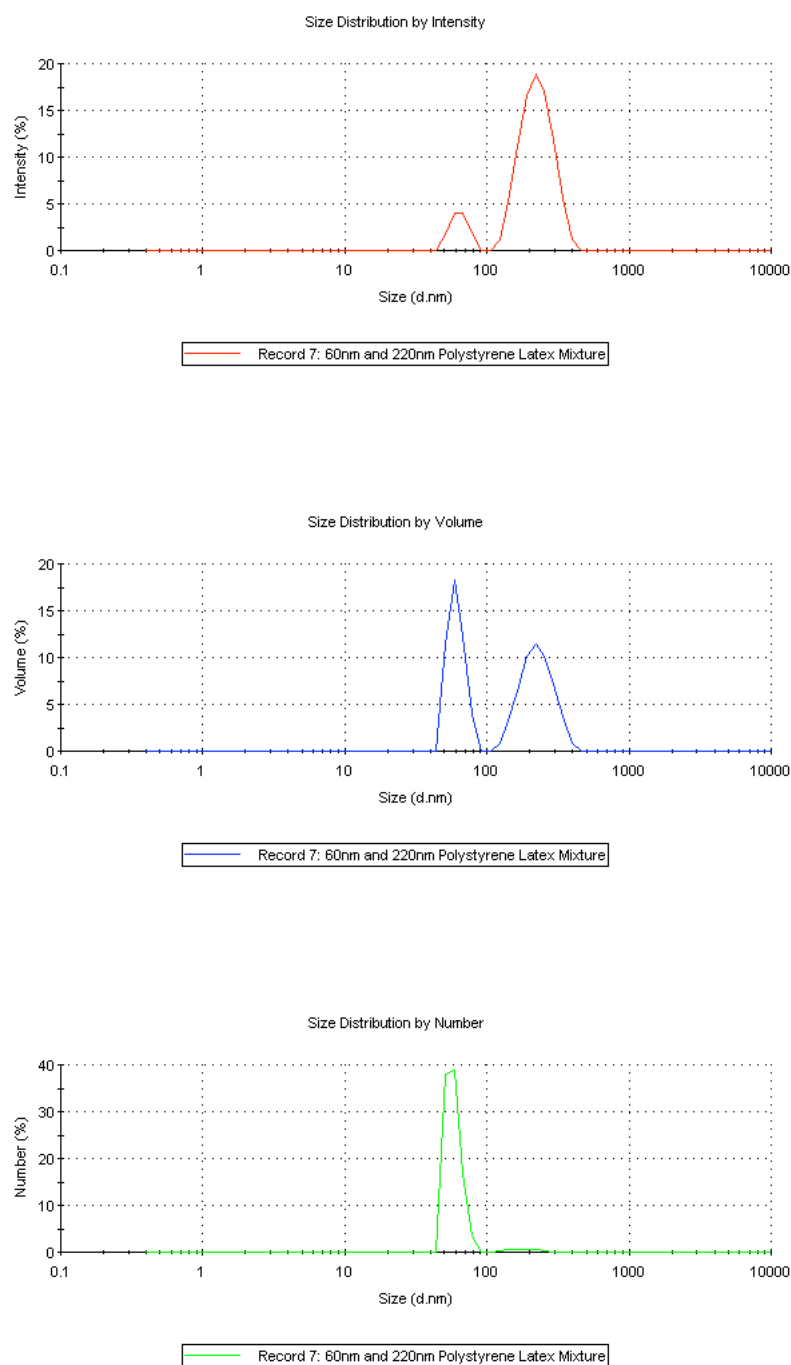


Figure 2.3: Size distribution by intensity, volume and number of a mixture of 60nm and 220nm latex beads. The larger beads scatter more strongly in proportion to their size, so the peak at around 220nm is much larger than the peak at 60nm in the intensity distribution (top). In the volume distribution (middle), the peaks are of approximately equal size, showing that the particles occupy similar volumes. Converting from a distribution by intensity to a volume distribution assumes that the particles are spherical. When this is converted into a distribution by number (bottom), the peak at 60nm dominates, while the peak at 220nm almost disappears. This indicates that there are many more of the 60nm particles than the 220nm particles. Since the initial distribution is based on intensity, any errors are compounded by subsequent transformations.

2.3.3.3 Application to protein crystallisation

Perhaps the most interesting application of dynamic light scattering from the point of view of an X-ray crystallographer is in gaining insights into the behaviour of proteins under various conditions related to their crystallisation. This may range from determining which buffers provide greatest stability for storage of precious protein samples, to confirming the results of a purification or even identifying optimum conditions for crystal growth.

Apart from the obviously useful ability to measure sample polydispersity and aggregation, DLS also allows several other factors relevant to protein crystallisation to be determined. The second virial coefficient is a thermodynamic property that describes the strength of the interaction between protein and solvent. This quality is linked with the tendency of a protein to either precipitate as amorphous aggregates or form ordered crystals at saturating concentrations. Thus DLS offers both a way of optimising crystallisation strategies and a powerful complement to the other techniques outlined in this chapter used to examine native protein structures.

2.3.4 Circular dichroism spectroscopy

Another highly useful technique for probing the solution structure of proteins is circular dichroism (CD) spectroscopy, which exploits the differential absorption of left- and right-handed circularly polarised light by chiral compounds. In the case of proteins, the chromophores that give rise to CD signals include the peptide bond, aromatic side chains and disulphide bonds, while associated cofactors and ligands bound in an asymmetric environment may also contribute to the CD spectrum.

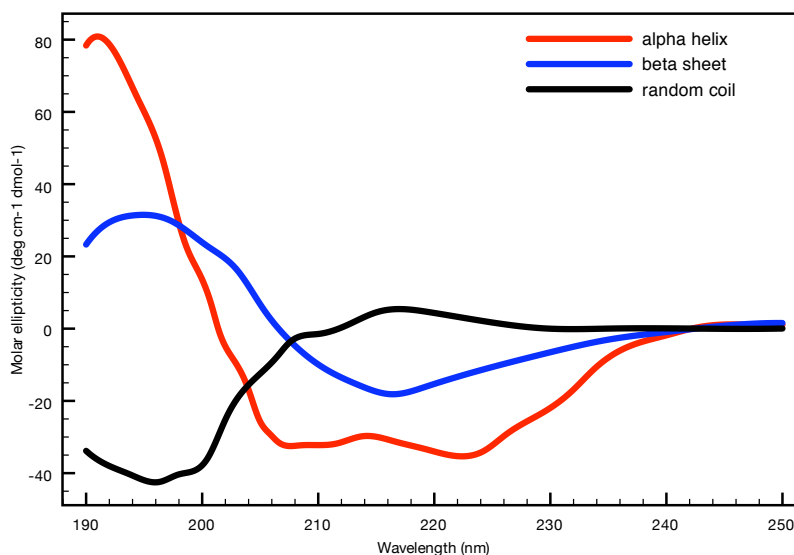


Figure 2.4: Characteristic CD spectra of various secondary structure conformations of the polypeptide chain. An α -helical structure gives a spectrum with two minima at around 210 and 220 nm, while β -structure displays a single minimum at around 215 nm. Peptides in a random coil conformation show an inversion of the spectrum.

In the far UV region, below 240 nm, absorption is primarily due to $n \rightarrow \pi^*$ (200 nm) and $\pi \rightarrow \pi^*$ (190 nm) transitions originating in the peptide bond. The regularly repeating peptide bond angles of protein secondary structural elements give rise to characteristic spectra in this region, as shown in Figure 2.4. A protein rich in alpha-

helical secondary structure displays minima at 208 and 222nm in the CD spectrum, while beta structure gives rise to a single minimum at 216nm and random coils show an inverted signal. It is thus possible to deconvolute the signal from a given protein to obtain an estimate of the secondary structure composition, by comparing experimental results with reference data from proteins whose structure has been solved by NMR or X-ray crystallography.

Further structural information can be obtained from near-UV region spectra, at 260–300nm. In this region, absorption is due to aromatic side chains, giving a unique fingerprint of the tertiary structure of the protein. This information can be useful when comparing wild-type and mutant forms or can give evidence of a “molten globule” state. The near-UV CD spectrum can also be sensitive to small changes in tertiary structure resulting from ligand binding or changes in solvent conditions. The signal strength in the near-UV region is much weaker than in the far-UV, so these studies require higher sample concentrations and larger amounts.

During a CD experiment, the difference in absorption between the left- and right-handed circularly polarised beams is typically recorded as a function of wavelength and the results are commonly expressed as the molar ellipticity of the sample, as given by the following equation:

$$[\theta] = \frac{\theta \times M}{c \times l \times 10}$$

Equation 1: Molar ellipticity, $[\theta]$, is expressed in degrees $\text{cm}^2 \text{dmol}^{-1}$. The experimental values are typically recorded as ellipticity, θ . M is the relative molecular mass of the protein being examined, c is the concentration in g/ml and l is the path length of the CD cell in centimetres.

While CD offers a powerful insight into the solution structure of proteins, it is limited in that only an overall signal can be observed. Furthermore, obtaining estimates of secondary structure rely on extremely accurate experimental conditions and the availability of suitable reference data. However, CD constitutes a valuable tool for tracking protein conformational changes, particularly where large changes in secondary structure occur, for example in response to ligand binding or denaturation.

2.3.5 Transmission electron microscopy

Moving away from studying proteins in solution, we come to the use of transmission electron microscopy (TEM) in the context of structural biology. As outlined in section 2.5, electron microscopy can be used to resolve much finer detail in a sample than light microscopy and is therefore another valuable tool with which we can study protein structures. There are several different types of electron microscopy, of which only TEM will be discussed here, it being the most applicable to the samples and aims in question.

As the name suggests, in TEM the electrons are transmitted through the sample, much like light is passed through a thin specimen in light microscopy. A beam of electrons is used as the illuminating source and electromagnetic lenses replace the traditional glass lenses used in a light microscope. In addition, the entire assembly is maintained under vacuum: firstly to prevent electric arcing due to the voltage difference between the cathode and the ground; and secondly to increase the mean free path of the electrons within the instrument, thereby allowing the electron beam to reach the sample without being attenuated by collisions with gas molecules. Instead of using an eyepiece, the image from a transmission electron microscope is displayed using projector lenses. These focus electrons, which cannot be detected by the human eye, onto a phosphor-coated screen that produces a fluorescent image where the electrons strike.

To prepare a sample for TEM, it must be thin enough so that the electrons can pass through it. In the case of biological samples, an additional factor to consider is that the sample must be able to withstand the vacuum inside the instrument, which can be achieved by fixing the samples. Depending on the nature of the sample and the purpose of the experiment, this may be done by embedding material such as cells or tissue in a plastic resin from which thin slices can then be cut using a microtome. Once the sample has been prepared, it is deposited onto a fine metal mesh or grid, which allows easier manipulation and handling. An electron lucent film, such as a thin layer of carbon, may be deposited upon the grid to aid in supporting the specimen.

Protein particles or other solvated or suspended samples are generally placed directly onto carbon-coated grids and may be fixated using chemical cross-linking. In most cases, additional staining is required to reveal the ultra-structure of a specimen, since many biological samples are virtually transparent to electrons. This typically involves negative staining, which uses heavy metals to increase the contrast of the TEM image. When a staining solution (e.g. uranyl acetate) is applied to the sample, the specimen is revealed as a bright feature against a dark background under the TEM, due to the scattering of the imaging electrons by the heavy atoms in the stain. Since the stain is able to penetrate the sample, rather than just coating it, this allows the fine structural detail of the sample to be observed.

Cryo-electron microscopy, on the other hand, allows the observation of samples without requiring modifications such as fixing or staining. Instead, the sample is rapidly frozen in liquid ethane and maintained at cryogenic temperatures throughout the experiment. In the case of proteins, this means that their native state is preserved, since the water in the sample freezes in a vitreous state and the sample in its solid state is protected from the vacuum inside the instrument.

For single particles such as protein structures, hundreds and thousands of individual molecules — each in a different orientation — can be recorded. A three-dimensional reconstruction can then be assembled from these series of two-dimensional images using an imaging strategy known as cryo-electron tomography (CET). This technique is a powerful complement to high-resolution techniques: for example, the structures of insoluble protein assemblies — which otherwise could not be crystallised or analysed by NMR — can be modelled by fitting high-resolution fragments that have been solved into a molecular envelope generated by CET.

Electron microscopy is a powerful tool in the structural biologist's arsenal. It allows the direct visualisation of macromolecular assemblies, and low-resolution three-dimensional structural information from cryo-electron microscopy may be invaluable in cases where other structural modelling techniques fail. In combination with high-

resolution structures from X-ray crystallography or NMR studies of proteins and protein fragments, detailed models of macromolecular assemblies can be constructed, which would not be possible using the individual techniques in isolation.

2.4 Protein crystallisation

When embarking on the structural characterisation of a protein by X-ray crystallography, the crucial and often the most time-consuming step is the production of suitable protein crystals. For crystallisation to be successful, a protein must be coaxed to precipitate into the solid phase in such a way that structural integrity is maintained and the appropriate crystal contacts can be made that allow crystal growth to occur. Given the nature of protein molecules when compared with small-molecular-mass compounds, this is no mean feat.

2.4.1 Crystallisation theory

In contrast to small molecules, which are often crystallised by simple evaporation of the solvent to dryness, protein crystals are grown almost exclusively from supersaturated solutions. This is a consequence of the chemical environment in which proteins have evolved, which is characterised by aqueous chemistry within the narrow range of temperature and pH typical of biological systems (McPherson, 2004). Although the strategies employed in protein crystallisation experiments have been extensively refined over the years and are now both much more diverse and rational, the basic technique remains relatively unchanged.

Essentially, all protein crystallisation trials involve an empirical search for suitable conditions using a combination of variables: protein concentration, temperature, pH, precipitants such as salts or PEGs, and various other additives such as cofactors and metal ions. Owing to the number of parameters involved, the initial screening for suitable crystallisation conditions is generally performed in multi-well plates, either manually or increasingly using robotics. During the experiment, the concentration of protein and/or precipitant in solution is gradually varied, either by vapour diffusion, dialysis, or free interface diffusion, with the aim of achieving a state of protein supersaturation that will induce nucleation and crystal growth (Chayen, 2004). A typical example of a crystallisation experiment is given in the phase diagram in Figure 2.5.

At the beginning of the experiment, the conditions maintain the protein in soluble form, but as water is removed from the crystallisation solution a state of supersaturation is attained. Crystals grow from nuclei composed of a few ordered molecules, and therefore a successful crystallisation experiment must induce the protein to pass through conditions that promote nucleation as opposed to random aggregation. As the crystals grow, the concentration of protein in the crystallisation solution

decreases until an equilibrium is reached. An ideal experiment will produce few nuclei, thereby maximising the amount of protein available to each growing crystal and resulting in a handful of large crystals.

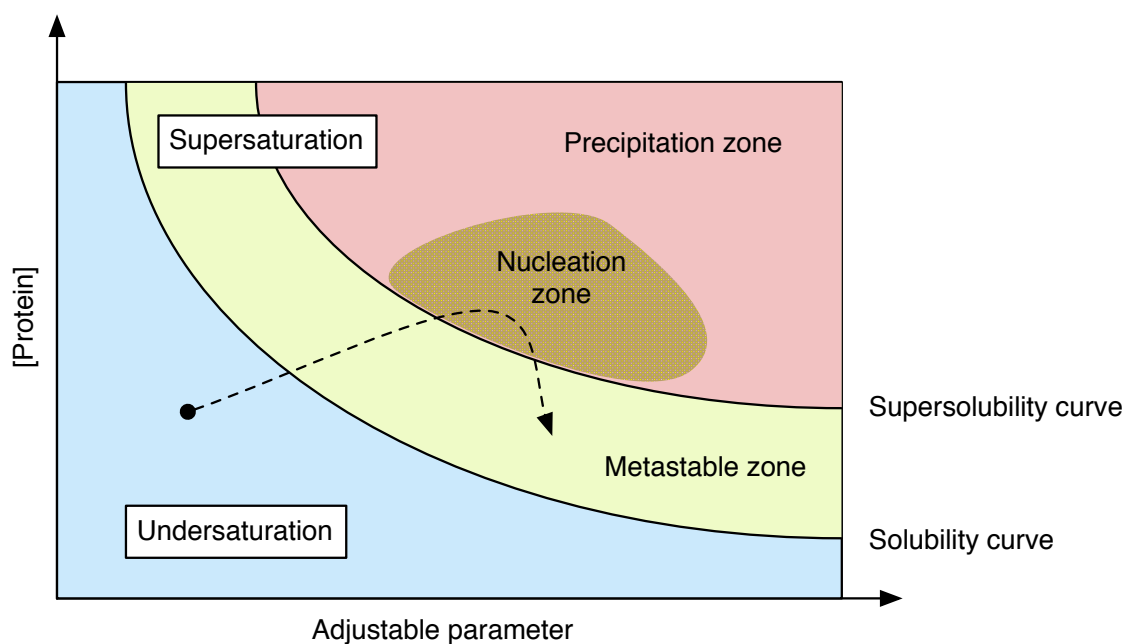


Figure 2.5: Illustration of a typical protein phase diagram. The adjustable parameter could be temperature, pH, precipitant concentration etc. The dashed line shows the path taken through the various regions of the phase diagram during a successful experiment, with the starting conditions represented by a black circle.

The path taken by a protein through crystallisation space is affected by innumerable factors, and the complexity of this process is why crystallisation is often referred to as a black art. Nonetheless, the effects of precipitants on proteins in solution can be broadly distinguished and the kinetics of a crystallisation experiment can be understood in terms of the various stages of crystal growth. These two key areas will be explored in the following sections.

2.4.1.1 *Precipitant effects*

Precipitants can be classified as salts, long chain polymers, low molecular weight polymers and non-volatile organic compounds, and organic solvents. These various precipitant classes differ in the mechanisms by which they induce proteins to crystallise. Given the connate dependence of protein structures on their environment, and their complexity in terms of the interactions both intramolecularly and with the surrounding solvent, it is perhaps surprising that trends in the mode of action of precipitants can be distinguished. Nonetheless, the electrostatic, hydrophilic/hydrophobic and steric properties of the various precipitant classes afford some insight into their general effects on protein structure and solubility. Of course, these generalisations apply to typical proteins — those of intermediate molecular weight, with a globular structure and a net negative charge at neutral pH — whereas the effects on proteins which possess unusual physicochemical characteristics are likely to differ accordingly.

a) Neutral salts and the Hofmeister series

Neutral salts primarily promote protein crystallisation by competing for the binding of water molecules. High concentrations of neutral salts cause solvent-deprived macromolecules to aggregate either as amorphous precipitates or as ordered crystals; a phenomenon commonly referred to as “salting out”. The most efficient protein precipitants are multivalent anions, such as sulphates, phosphates and citrates, with ammonium sulphate typifying precipitants in this class (McPherson, 2004). Small ions with a high charge density like these effectively strengthen the hydrophobic effect by enhancing the structure of bulk water, which consists of a quasi-crystalline network of hydrogen bonds. Since they increase the order of water, such ions are known as kosmotropes.

Anions and cations can be arranged in order of their ability to modify the structure of water and thereby affect protein solubility, a classification known as the Hofmeister series¹. Strongly hydrated, kosmotropic ions early in the series exert a stabilising effect on proteins, while chaotropic ions at the end of the series disrupt water structure, weakening the hydrophobic effect and thereby increasing the solvent accessible surface of the protein. This has the effect of destabilising protein structure, but also promoting solubility (Collins, 2004). In addition to their effects on bulk water, neutral salts affect water molecules in hydration shells around protein surfaces in a similar manner, due to preferential interactions between solvent, solute and protein molecules. Aside from the effects on aqueous hydrogen bonding, ions and other solutes also interact specifically with protein moieties such as the peptide backbone or side chains, adding further complexity to the interactions governing protein crystallisation.

b) Long chain polymers

In contrast to neutral salts, where electrostatic forces dominate interactions between the solute, solvent and protein in solution, long chain organic polymers precipitate proteins primarily through excluded volume effects. Here, the underlying principle is that long chain polymers, which unlike proteins have no defined solution structure, occupy a certain volume that cannot be simultaneously occupied by any other species. As a result, there is less solvent space accessible to any protein molecules in solution, forcing them to aggregate and finally transition into the solid state. By far the most common long chain polymers used as protein crystallisation agents are polyethylene glycols, a family of hydrophilic nonionic polyether compounds with a wide variety of uses in industry and medicine.

1. Not to be confused with the Hoffmeister series, otherwise known as Baywatch.

Like the ionic precipitants discussed previously, PEGs differ in their abilities to force proteins from solution, with higher molecular weight polymers possessing a greater capacity to drive the crystallisation process thanks to their larger excluded volume. However, the molecular weight of the protein in question bears no relation to the size of the polymer best used for its crystallisation, and indeed PEGs with similar molecular weights are not necessarily interchangeable for a given protein (McPherson, 2004). This means that the search for a PEG or other organic polymer to act as a suitable protein precipitant is unavoidably an empirical process.

c) Low molecular weight organic compounds

This class of precipitants includes low molecular weight (< 1000) PEGs, MPD and other small, non-volatile organic compounds. The modes of action of this class of precipitants are variable and not entirely understood. MPD appears to be preferentially excluded from protein-solvent interfaces due to repulsion by surface charges, while for PEGs excluded volume effects remain dominant. Other precipitants in this class, which are commonly used as additives to supplement another major precipitant, are various naturally occurring osmolytes such as sugars and amino acids (Bolen, 2004). As with other precipitant classes, the interplay between solvophobic/solvophilic effects, steric factors and specific interactions with protein moieties governs the overall impact of these precipitants on the crystallisation process.

d) Organic solvents

Organic solvents offer yet another means of manipulating the solution conditions of a protein crystallisation experiment by reducing the dielectric of the surrounding medium. The introduction of non-polar solvents reduces the activity coefficient of water, making it a poorer solvent and thus pushing solvated proteins towards the solid state (Cohn et al., 1947). Again, specific interactions with portions of the protein such as hydrophobic side chains modulate these general effects. Low concentrations of certain solvents can act to stabilise protein structures while high concentrations act as denaturants and others have little or no effect on protein structural stability (Asakura et al., 1978). The volatility of organic solvents means that any crystallisation experiment in which they are employed must be carefully designed to take this into account.

2.4.1.2 *Protein crystal growth*

Protein crystallisation proceeds via distinct stages. These can be distinguished as nucleation, growth and finally the cessation of growth. Different processes govern each stage and they are all affected by different factors during a protein crystallisation experiment. These various mechanisms and the practical considerations that were applicable to the work of this thesis will be briefly explored in the following paragraphs.

a) Nucleation

The first crucial event during the crystallisation of a biological macromolecule is nucleation. This process is analogous to the condensation of droplets of steam on a bathroom mirror. A comprehensive review of the nucleation of protein crystals can be found elsewhere (Manuel Garcia-Ruiz, 2003) but a short summary of the process will be given here, along with experimental aspects. Thermodynamics dictates that a supersaturated solution must return to equilibrium by segregating a solid phase until equilibrium is established. However, in practice the threshold value of supersaturation must be exceeded for precipitation to occur spontaneously. This paradox is resolved when nucleation is taken into account.

Under normal crystallisation conditions, the free energy of a crystalline aggregate follows a curved trajectory as it increases in size, first increasing to a maximum ΔG and decreasing steadily thereafter. In contrast, the formation of disordered precipitate follows a shallower path of decreasing ΔG with a much lower energy barrier. The height of this initial increase in energy dictates the time required for crystal nuclei to form, and depends strongly on the supersaturation of the protein solution. Once a crystal nucleus has formed, it grows slowly at first, since attachment and disassociation of individual molecules occur with almost equal probabilities. This process is governed by the equilibrium rate constants for monomer-monomer and

monomer-crystal association. These two rate constants, along with the concentration of monomers in solution, also determine the critical number of molecules required to push the growing aggregate through its energy maximum (Durbin and Feher, 1996). Of course, apart from the thermodynamic elements outlined above, the outcome of any crystallisation experiment is also dictated by kinetic factors.

Understanding nucleation will not help us to avoid the necessary exploration of the phase landscape of protein systems in the search for crystals and their atomic structures. But a better general understanding of the nucleation process can aid in controlling the experimental variables implicated in nucleation according to our current understanding. The nucleation step implicitly controls the structure of the crystallising phase and the number of particles (and thus the crystal size) appearing in a crystallisation system (Manuel Garcia-Ruiz, 2003)

b) Growth

When a crystal is born from its supersaturated mother liquor, it will grow until it reaches an equilibrium with the concentration of protein in solution. Both kinetic and thermodynamic factors affect this process, and these are detailed by Chernov, 2003. While the nucleation processes just discussed typically require high protein concentrations, crystal growth proceeds better in a region of the phase diagram termed the metastable zone. In this region, nucleation does not occur, but crystal growth is able to proceed unhindered until either the equilibrium concentration is reached or the crystal growth terminates for other reasons.

c) Order and disorder in protein crystals

While we typically think of crystals as idealised structures, in reality they are imperfect. Often, the larger a crystal grows the more defects it will accumulate, although the physical appearance of a crystal often has no bearing on the diffraction quality. Of course, thanks to synchrotron sources becoming more commonly available, as well as technological innovations in crystal handling and mounting methodology, it is now possible to work with smaller crystals than ever before. Nonetheless, the aim of any crystallisation experiment is to produce the best quality diffraction data possible, and this requires well-ordered protein crystals.

The aim of X-ray crystallography is to determine the positions of atoms within a macromolecule and it relies on the cumulative signal from thousands of atoms arranged identically in each molecule occupying the crystal lattice. However, the proteins constituting a real crystal may not all be perfectly aligned. There may be conventional defects of the crystal lattice: point defects, which occur at a single position in the crystal lattice; stacking faults, where the regular packing of crystal layers is interrupted; or mosaic blocks that cause different parts of the crystal to be slightly misaligned. These types of disorder can usually be detected by the mosaicity. Crystals may also exhibit twinning, which occurs when two separate crystal domains are conjoined in such a way that their diffraction lattices overlap, either partially or completely. Twinning also affects the mosaicity of diffraction data and can be detected by various other criteria as discussed in detail in section 6.3.3, p. 270.

In addition to the relatively macroscopic defects already mentioned, disturbances in the crystal structure may arise through different conformations of the molecules themselves. For example, the protein in question may be an enzyme trapped in various states of catalysis, or possess flexible hinge regions. While disorder in protein crystals is generally undesirable, this type of pervasive, localised variation mainly affects the atomic B-factors of relevant parts of the protein structure, and may offer some insight into the function of the macromolecule.

2.4.2 Crystallisation methodology

A relatively common set-up for crystallisation experiments, and the one used primarily in the work presented here, is the hanging drop method. This involves placing a small volume — usually between 1 and 10 μl — of protein solution on a small glass or plastic slide (such as a microscopy cover slip) and mixing it with an equal volume of the test solution. The cover slip is inverted and placed over a large reservoir of the test solution, using grease to form an airtight seal, and the solutions are then left to equilibrate. As water evaporates from the drop, the concentration of both the protein and the precipitant increases until the water concentration in the drop matches that of the reservoir. The drop is then observed at regular intervals using a microscope and the results are recorded. Protein crystals can form in this mother liquor during any time from minutes to years from the set up of the experiment, although eventually all the water escapes from the system and the drop evaporates to dryness, destroying any protein in the process.

2.4.2.1 *Controlling nucleation and growth*

Nucleation can be spontaneous (as described above) or may result from the introduction of seeds or nucleants into the crystallisation solution. Seeds are typically nuclei transferred by crushing or scraping crystals of the same protein that have already grown, but are perhaps of lesser quality. Nucleants on the other hand, are unrelated substances that have nucleation inducing properties. These may be specific minerals or porous substances introduced as a variable in the experiment, or simply bits of dust that have ended up in the crystallisation drop. They promote crystal growth by providing sites for nuclei to form and result in epitaxial growth of protein crystals from the nucleant (Chayen et al., 2006). By including seeding as a parameter in crystallisation trials, we can control the number and size of any crystals.

However, there are further factors to be considered. The surfaces of crystals growing from impure samples may become poisoned and stop growing, possibly also resulting in an increase in the disorder of the crystal. Furthermore, the protein concentration must be just right to allow large crystals to develop. If the protein concentration is too high, many nuclei may form simultaneously. Since each will consume protein at approximately the same rate, this will result in the growth of hundreds of tiny crystals. On the other hand, at low protein concentrations, where only a few nuclei have formed, an equilibrium may be reached between the crystal and aqueous phases before the crystals have developed to a reasonable size.

Equally, the quality and size of macromolecular crystals can be varied by changing the parameters involved in controlling the equilibration of the drop. This can be affected by factors such as temperature, protein and precipitant concentration, and the size of the drops. Common temperatures for performing crystallisation experiments are 4°C and room temperature, although some proteins require treatment at higher temperatures for successful crystallisation (Cudney, 1999). Temperature also influences protein and precipitant solubility, and can affect nucleation rates. Drop size, or more specifically the surface-area-to-volume ratio, is also important when it comes to controlling the rate of equilibration in an experiment.

2.4.2.2 Interpreting the results of a crystallisation experiment

The outcome of a crystallisation experiment can be any one of several possibilities. In the ideal case, the result will be one or several large, well-ordered crystals. However, in practice, this is a rare occurrence. If the path through crystallisation space has remained primarily in the undersaturated region of the phase diagram, then the result will be a clear drop. If the concentration of either protein or precipitant was too high, however, then an amorphous precipitate of denatured protein will be observed. This may also be the case if the protein is incompatible with any of the substances used in

the test solution, or because of extremes in temperature and pH. Successful crystallisation depends on hitting the ‘sweet spot’ in terms of experimental conditions and kinetics that pass through the nucleation zone and into the metastable zone allowing crystal growth to occur, while maintaining the structural integrity of the protein. In order to assess the results of crystallisation trials, the scoring system outlined in Table 2 was used.


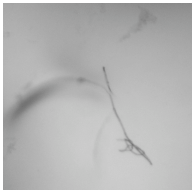

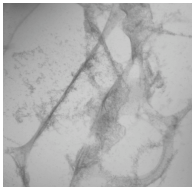
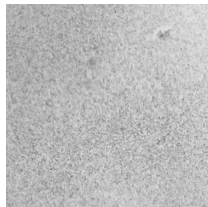
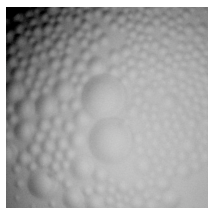
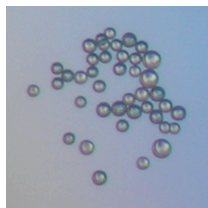

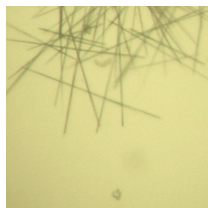
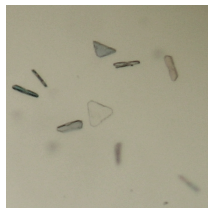
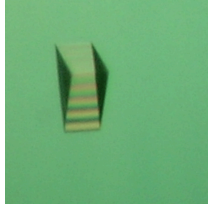
Score	Drop description	Example
<i>Conditions unlikely to yield crystals</i>		
0	Clear	
1	Contains non-protein particles	
2	Light precipitate	
3	Heavy amorphous precipitate	

Table 2: Scores used in assessing the results of crystallisation experiments. The scores are arbitrary, but can be grouped to reflect conditions that are either far from optimal, or close to the conditions required for crystallisation. An example of the observed drop that corresponds to each score is shown.

Score	Drop description	Example
<i>Crystallisation conditions close to optimal</i>		
4	Gelatinous precipitate	
5	Phase separation	
6	Spherulites	
7	Microcrystals	
8	Needles (1D crystals)	
9	Plates (2D crystals)	
10	Crystals (3D crystals)	

2.4.2.3 *Determining whether a crystal is protein or salt*

Protein crystals display a staggering range of morphologies, and a trained eye is required to discern potential crystals from the debris and other matter that may accumulate in the crystallisation drop. Identification of protein crystals can be aided by various methods, such as using polarised light to test for birefringence. However, given that protein crystallisation experiments can often also induce the crystallisation of salts or other substances present in the test solution, observing a birefringent object in the crystallisation drop is not always indicative of a protein crystal.

A more accurate assessment of the nature of any crystalline materials can be made using more destructive methods. In the ‘crunch test’, a putative protein crystal is crushed using a fine needle. If the crystal is resistant to destruction, or crushes with a palpable ‘crunch’, then it is more than likely a salt crystal. Protein crystals, on the other hand, are relatively soft and correspondingly fragile. This is primarily due to the low density of bonds within the crystal contacts of macromolecular crystals and the small size of the contact areas when compared with the whole protein surface (Chernov, 2003).

The delicate nature of protein crystals, which is also in part attributed to a relatively high solvent content of between 25–90% (McPherson, 2004), can be turned to an advantage in another test. In this test, the small-molecule dye Izit (Hampton Research, USA) is added to the crystallisation drop, where it can penetrate the solvent channels in macromolecular crystals but is excluded from small-molecule and inorganic crystals. This causes any protein crystals present in the drop to take on a blue colour, allowing them to be easily identified. Unfortunately, the test is likely to interfere with any further work on the crystals.

The ultimate test of both the nature and quality of any crystalline substances in the crystallisation drop is, of course, the diffraction experiment itself. Protein crystals are instantly recognisable by their diffraction patterns, which typically feature closely

spaced spots due to the relatively large unit cell dimensions of macromolecular crystals. Diffraction theory and the diffraction of protein crystals are covered more extensively in section 2.5.2, p. 98.

2.4.2.4 Optimising crystallisation conditions

Often, the initial rounds of screening will yield conditions that promote crystallisation but result in sub-optimal crystals. These may take various forms, such as hundreds of microcrystals due to excessive nucleation; thin fragile needles and plates as a result of rapid growth along one or two axes; or spherical crystalline bodies known as spherulites. The conditions yielding these initial ‘hits’ can be expanded into further screens that sample crystallisation space at much finer intervals.

Even if the initial screening did not yield crystals, similar optimisations can be performed around conditions that yielded light, granular precipitates, or where a small change in one of the parameters of the experiment resulted in a transition from a clear drop to precipitate. This is because conditions such as these are likely to be located near the metastable zone on the crystallisation phase diagram (see Figure 2.5, p. 75), and so exploring the crystallisation space around them might result in successful crystallisation.

In addition to simply varying the experimental parameters to achieve optimal nucleation and growth conditions, protein crystallographers can make use of seeding techniques to produce the large, single crystals desired for diffraction purposes. These methods involve transferring crystal nuclei from sub-optimal crystalline entities into conditions likely to be located in the metastable zone. In this way, excessive nucleation and rapid growth are prevented, and the results are likely to be more desirable in terms of the number and size of crystals produced. In the case of too little nucleation, external nucleants can be introduced to provide impurities that can act as sites for

nucleation and growth. Furthermore, the kinetics of the experiment can be varied by changing the drop size (surface area to volume ratio), temperature or other factors that may affect the rate of equilibration.

While in theory it would be preferable to perform an extended series of crystallisation experiments in order to produce optimal crystals for diffraction, in practice such large-scale trials are generally limited by the amount of protein material available. This is primarily because of the high concentrations typically required for crystallisation — often tens of mg/ml — and the arbitrary nature of the experiments. When commencing a new crystallisation strategy, a crystallographer can use prior knowledge about the protein in question, together with solubility information gleaned from systematic initial screening or other methods, to tailor the strategy in order to conserve material and maximise the chances of obtaining suitable crystals. However, sometimes the appearance of protein crystals is a matter of sheer luck. Something as seemingly trivial as whether the protein sample is kept on ice or held in the experimenter's hand while setting up the drops can mean the difference between crystals and no crystals (Cudney, 1999). Detailed laboratory records are therefore paramount.

The results of crystallisation trials described in this work were typically evaluated once a day for the first week, and monthly thereafter until the drops dried up completely or no further changes were observed. The results of the experiments were recorded by observing each drop under a microscope using bright-field and cross-polarised illumination. Each drop was assigned an arbitrary score, ranging from 0 for a clear drop to 10 for three-dimensional crystals, as previously outlined in Table 2, p. 85. This offers a way of quantitating the data and allows different crystallisation experiments to be compared. Obtaining crystals of a protein can be a daunting task, but is an absolutely necessary undertaking in order to unlock the protein structure contained within. While NMR is useful for smaller proteins, larger molecules can give

crowded spectra with overlapping peaks, making them difficult to interpret. Crystallography therefore remains the established approach for determining high-resolution structures of large protein assemblies.

2.5 X-ray crystallography

As already mentioned, X-ray crystallography is one of two key techniques commonly used to obtain information about the structure of proteins at atomic resolution, the other technique being NMR. Crystallographic analyses are applicable to a much broader range of proteins, since there is no limit to the size of the molecule under scrutiny; the only real requirement is that the protein must form crystals from which useful data can be collected.

In order to obtain information about a protein at the atomic scale, a suitable probe must be employed. Visible light has a wavelength in the range of hundreds of nanometres, making light microscopy a good tool with which to probe samples of that order of size and greater, such as cells and indeed the macrostructure of protein crystals. However, if we wish to resolve finer detail, we need a correspondingly finer probe. Electron microscopy makes use of electrons, which have a smaller wavelength than visible light, to probe cellular ultra-structure and large protein assemblies, but it cannot give information at the level of individual atoms. It so happens that X-rays have wavelengths that correspond roughly to the diameter of individual atoms, at around one Angstrom, making them a suitable probe for structural information at the atomic scale.

Since X-ray wavelengths are so much smaller than those of visible light and electrons, they also have correspondingly higher energies. This means that they interact with matter in different ways, posing a great challenge to the crystallographer. While light and electron microscopy rely on lenses to create an image of the sample from the scattered radiation, it is impossible to build a lens that will focus X-rays. Instead, the image must be constructed computationally from the diffraction data, as will be explained below. The following sections explain the individual steps between successful crystallisation of a protein and the creation of a structural model at atomic resolution based on the available information.

2.5.1 Principles of diffraction

The key principle behind X-ray crystallography is the scattering of X-rays by the electrons of molecules contained within the unit cell, the basic repeating unit that makes up the crystal structure by translation alone. In three-dimensional crystals, the unit cell can be described by three vectors (**a**, **b** and **c**) and three angles between them (α , β and γ). Within the unit cell, individual asymmetric units may be related by crystallographic symmetry, so that more than one identical molecule can occupy the unit cell (see Figure 2.6). Of course, the actual number of molecules in the unit cell is not necessarily equal to the number of asymmetric units dictated by the crystallographic symmetry.

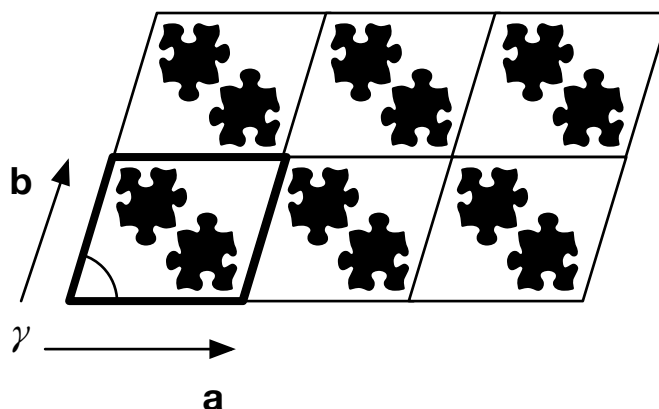


Figure 2.6: Illustration of a simple protein crystal in two dimensions. The unit cell (outlined in bold) can be described by the vectors **a** and **b**, which run along the x and y axes, and the angle between them, γ . The unit cell can be repeatedly translated in all directions to map out the crystal structure. Within a single unit cell there can be one or several protein molecules. Here, two copies of the protein are shown, related by a 180° rotation around a symmetry axis perpendicular to the plane of the figure.

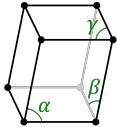
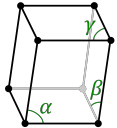
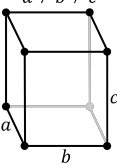
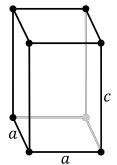
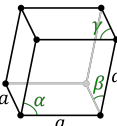
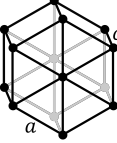
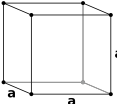
Crystal system	Required symmetry	Unit cell parameters
Triclinic	None	$\alpha, \beta, \gamma \neq 90^\circ$ 
Monoclinic	2-fold	$\alpha \neq 90^\circ$ $\beta, \gamma = 90^\circ$ 
Orthorhombic	3 x 2-fold	$a \neq b \neq c$ 
Tetragonal	4-fold	$a \neq c$ 
Trigonal	3-fold	$\alpha = \beta = \gamma \neq 90^\circ$ 
Hexagonal	6-fold	
Cubic	4 x 3-fold	

Table 3: The seven crystal systems. The required symmetry of the space groups and the unit cell parameters corresponding to each system are shown.

The repeating nature of the crystal means a large number of molecules are arranged in virtually identical orientations. In other words, we have a periodic array of scatterers. When a crystal is placed in an X-ray beam, the incident rays are re-emitted in all directions by the electrons they interact with. Under certain conditions, these interfere constructively, producing an intense signal that can be recorded as the characteristic spots of a diffraction pattern. The positions of the spots are determined by the unit cell parameters and the intrinsic symmetry of the crystal, while the intensities of the spots are determined by the electron density within the crystal. The electron density within each unit cell can be described by a mathematical function and it is this function that we wish to determine in order to model the protein structure.

2.5.1.1 Bragg's law and reciprocal space

Diffraction can be described in terms of the reflection of incoming rays from a set of hypothetical parallel planes that bisect the crystal. If a detector is placed to one side of the crystal, the scattered rays will all travel different distances before they encounter the detector (see Figure 2.7). Constructive interference of scattered rays — resulting in a measurable signal — occurs when the difference between their path lengths is a multiple of the wavelength. This is because the scattered X-ray waves are in phase, that is, the peak of each wave is aligned with the peaks of all the other waves. These conditions, which result in measurable diffraction, are defined by Bragg's law:

$$n\lambda = 2d \sin \theta$$

Equation 2: Bragg's law, where n is an integer, λ is the wavelength of incident radiation, d is the distance between planes and θ is the scattering angle.

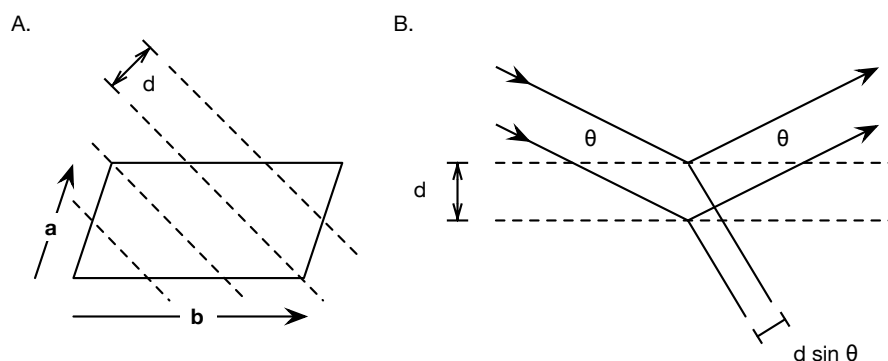


Figure 2.7: Bragg diffraction. A. To illustrate the concept of lattice planes a two-dimensional unit cell with axes **a** and **b** is shown. A set of hypothetical parallel planes, separated by distance d , are shown as dashed lines. The set of planes bisects the unit cell once along **a** and twice along **b**, giving the indices (1,2) for this set. B. Two lattice planes (dashed lines) are shown separated by distance d . The incident and reflected beams are shown as solid lines, making an angle θ with the lattice planes. The path difference between the rays reflecting from the two different planes is $2d \sin \theta$.

Families of lattice planes, and their corresponding reflections, are defined by Miller indices, written $(h \ k \ l)$ in three dimensions. The index represents the number of times the set of planes bisects the unit cell along the corresponding axis. As the index number increases, the spacing between planes decreases. Atoms lying close to a set of planes will tend to scatter in phase, so the intensity of each reflection tells us something about the distribution of the atoms relative to the corresponding set of Bragg planes. If the amplitude is weak, then the atoms are positioned randomly with respect to that set of planes, while a strong reflection indicates that the atoms are concentrated parallel to the Bragg planes. In addition, reflections with higher Miller indices contain information about atoms that are close together in the unit cell, so the highest resolution data are found at high scattering angles. This is a consequence of the effect of reciprocal space.

A diffraction pattern is a representation of the real crystal lattice in reciprocal space. The relationship between real space and reciprocal space is essentially that points close in real space are far apart in reciprocal space. The symmetry of the reciprocal lattice follows the symmetry of the real space lattice, and also, in the absence of anomalous scattering, the diffracted intensities obey Friedel's law:

$$I(h\ k\ l) = I(-h\ -k\ -l)$$

Equation 3: Friedel's law. The intensity, I , of a reflection $(h\ k\ l)$, is equal to the intensity of the reflection $(-h\ -k\ -l)$. This means that the reciprocal lattice is symmetrical.

2.5.1.2 *Structure factors and Fourier transforms*

A structure factor is a complex number with amplitude and phase that represents the wave resulting from diffraction. For several electrons distributed in space, the structure factor becomes the sum of the individual scattering contribution of each electron. The electron density of a protein crystal is a continuous function, so the structure factor in this case can be expressed as an integral. The structure factor and the electron density can thereby be shown to be related in a special way, known as a Fourier transform. It is this relationship that allows us to calculate a model of the electron density from the diffraction data of a protein crystal.

A Fourier transform is a mathematical operation that allows us to transform the variables of complex functions. A special property of the Fourier transform is that if it is applied to a function, and then applied again, the original function is restored. This allows us to transform the calculated structure factor (which is the Fourier transform of the electron density) into a model of the original electron density within the unit cell. However, the structure factor is a complex number and possesses both amplitude and phase. The phases of the reflections cannot be experimentally recorded, while the structure factors cannot be calculated using the intensity data alone. This forms the basis of the 'phase problem', which is described more fully in section 2.5.5, p. 109. The importance of phase information is shown in .



Figure 2.8: The importance of phase information. A. The images show the mathematician and physicist Joseph Fourier (left), and the pioneer of protein crystallography, Dorothy Hodgkin (right). B. The Fourier transform of the image of Joseph Fourier. From left to right: the logarithm of the amplitudes, the amplitude information only, and phase information only. C. An image reconstructed by combining the amplitudes of Fourier with the phases of Dorothy. The phases clearly contain more information about the image than the amplitudes, since we see Dorothy in the image. D. Images reconstructed by combining the phases of Fourier with two unrelated sets of amplitudes. Sets of amplitudes from two unrelated images (not shown) result in a similar reconstructed image. A good reconstruction requires both accurate phases and amplitudes to restore the original image.

2.5.2 The diffraction experiment

Once a protein crystal has been grown, the first step towards structure solution is to perform the diffraction experiment itself. This is generally carried out using an in-house X-ray source, in which X-rays are produced by accelerating electrons at high voltage against a metal target; or at a synchrotron facility, where radiation is emitted by electrons and positrons in high-energy storage rings. The process can generally be divided into two stages: the first being the preliminary characterisation of the crystal, to ascertain the maximal resolution to which it diffracts, the space group, and crystal quality; and the second being the collection of a usable data set (McRee, 1999).

The experimental set-up initially involves mounting the crystal on a goniometer placed in the path of the incident X-ray beam. A metal beam-stop placed behind the crystal prevents the direct beam from striking the detector, thus allowing the relatively weak diffracted rays to be recorded. There are various possible geometries for the diffraction experiment, depending on the instrumentation available, but here a rotation instrument will be described, which uses a two-dimensional detector to collect contiguous diffraction images generated by oscillating the crystal through a small angle. However, all cases involve rotation of the crystal within the incident beam such that the positions and intensities of all (or an appropriate number) of reflections can be measured. In order for this to be possible, all corresponding points on the reciprocal lattice must pass through the surface of the sphere with radius $1/\lambda$ (the Ewald sphere), thereby fulfilling the Bragg diffraction conditions, as illustrated in Figure 2.9.

An appropriate geometry is required for successful data collection, and structure solution relies heavily on accurate definition and recording of the parameters of the experiment. The crystal-to-detector distance must be adjusted so that the high-angle reflections — corresponding to the highest resolution to which the crystal diffracts — fall within range of the detector, if these are to be collected. Furthermore, the oscillation angle must be adjusted depending on the unit cell size, spot size, and resolution. In addition, the overall angle through which the crystal is rotated must be

determined, based on the symmetry of the crystal as gauged by the preliminary characterisation. Another consideration is the orientation of the crystal within the beam, relative to the detector. If the crystal is rotated around a single axis, there is a region of reciprocal space around the axis of rotation that can never be recorded, known as the blind region. This can be problematic if the crystal is oriented such that one of the crystal axes is along the rotation axis, as this will cause the resulting data to be incomplete. Overall, the data collection strategy involves careful selection of experimental parameters to yield the best possible data, and this is often achieved with the use of sophisticated software.

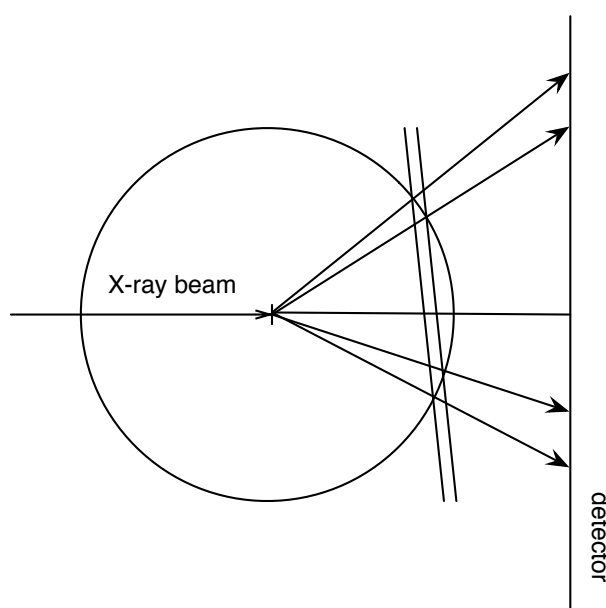


Figure 2.9: Diagrammatic representation of an X-ray diffraction experiment. The crystal is located at the centre of the sphere, whose dimensions are determined by the wavelength of the X-ray radiation. The diffraction pattern recorded on the detector can be regarded as the result of the intersection of reciprocal lattice points, located on a series of parallel planes, with the Ewald sphere. The dimensions of the reciprocal lattice are determined by the unit cell of the crystal.

2.5.3 Cryocrystallography

The diffraction experiment can be performed either at room temperature, or under cryogenic conditions. There are advantages and disadvantages to each method, and again the choice of which strategy to pursue depends on the particular characteristics of the crystal under study. Mounting a crystal for room temperature diffraction can be achieved by sealing the crystal in a thin capillary, surrounded by the mother liquor from which it was grown. The major advantage of room temperature experiments is that no unnecessary stresses are placed on the fragile crystals, apart from those to which it is exposed during handling, which can be minimised by skilful manipulation in the hands of the crystallographer. However, crystals at room temperature are particularly susceptible to radiation damage, since free radicals are able to diffuse freely along the solvent channels within the crystal lattice and destroy bonds within and between protein molecules. Prolonged exposure to the X-ray beam can result in visible destruction of the crystal as crystal contacts are degraded, meaning that the amount of data collected from a single crystal at room temperature can be severely limited.

The problem of radiation damage can be largely overcome by rapidly cooling the crystal to cryogenic temperatures (so that the water surrounding the crystal freezes as amorphous ice) and maintaining this low temperature state throughout the data collection procedure. At the ~ 100 K temperatures typically used, the crystal lifetime is usually prolonged by a factor of ~ 70 (Nave and Garman, 2005), thereby allowing a complete data set, or even multiple data sets, to be collected from a single crystal. This is a consequence of the dramatically reduced rates of atomic diffusion within the vitreous matrix, which means that any free radicals are likely to self-quench and broken bonds are likely to re-form, so maintaining the long-range order of the crystal. Furthermore, the molecular motion is greatly reduced by cryogenic freezing, which can result in the trapping of single or multiple thermodynamically favoured conformers, with resulting effects on the diffraction properties of the crystal (McRee, 1999).

The water surrounding the crystal must be vitrified to prevent the formation of ice crystals that may disrupt the protein crystal structure or obscure reflections in the diffraction pattern. This typically requires a combination of rapid cooling and the introduction of cryoprotectants, which reduce the formation of ice by disrupting the local order of water and preventing nucleation of ice crystals. Cryoprotectants can be divided into two classes: salts, which co-ordinate water molecules thereby reducing the number of water molecules free to make ordered intermolecular hydrogen bonds; and bifunctional molecules, which, being both hydrophilic and hydrophobic, can disrupt local order without necessarily co-ordinating large numbers of water molecules. Rapid cooling of the crystal is achieved by snap freezing it and the surrounding drop of liquid in a stream of cold gas, usually nitrogen, which is positioned so as to maintain cryogenic temperatures throughout the data collection procedure.

To mount the crystal for cryocrystallography, a fine fibre loop mounted on the end of a pin is used to scoop the crystal from the surrounding mother liquor. Cryoprotectants can be introduced — if necessary — by sweeping the crystal through a fresh drop of the well solution used for crystallisation with an appropriate volume of cryoprotectant added. While the advantages of cryocrystallography outlined above generally result in increased quality of the diffraction data obtained compared with room temperature experiments, the stresses placed on the crystal throughout the procedure can cause disruption to the crystalline lattice. These stresses may arise from the introduction of cryoprotectants and from changes in pH and in the dielectric constant of water upon cooling. Cross-linking the crystals prior to flash freezing can sometimes help to maintain their structural stability.

2.5.4 Data collection and processing

Once the crystal is mounted and frozen, if desired, a few test shots are typically made to determine the optimal geometry for the diffraction experiment. This involves adjusting the detector distance, exposure time and various other parameters to obtain the best quality data. The initial images can be processed to determine the unit cell dimensions, predicted space group, mosaicity and other features of the crystal, which will help in deciding on a data collection strategy.

2.5.4.1 General data processing procedure

An ideal set of diffraction images should be as complete as possible, meaning that the crystal must be rotated through an angle appropriate to its symmetry to allow all possible reflections to be detected. A low-symmetry space groups such as P1 would require the crystal to be rotated through a full 360° to collect a complete set of reflections, while a crystal with high-symmetry would only require a much smaller wedge of data to be collected. In addition, the exposure time and detector distance must be optimised to collect weak and high-resolution reflections, which fall at the edge of the detector, while not overexposing stronger reflections in the diffraction image. Other factors as diverse as radiation damage to the protein during the experiment or the allocation of time at synchrotron sources may be factors to consider when planning a data collection strategy.

Once a set of images has been collected, they can be processed. Nowadays, this procedure is greatly aided by a host of computational tools, but the overall principle is the same as for manual processing. First, reflections must be identified by comparing their intensity with the background. This is a relatively easy task for humans, and even computational methods of data processing may require user input to identify reflections in an otherwise automated process. Once the spacing of the reciprocal lattice has been

determined from the positions of the spots in the diffraction pattern, the unit cell dimensions can be determined. The symmetry and mosaic spread of the crystal can also be estimated.

The reflections are indexed, which involves assigning the Miller indices of each recorded reflection. This is typically done using an automated procedure, during which a likely space group may be assigned based on any symmetry present in the diffraction data, systematic absences of reflections as a result of screw axes, and the unit cell parameters. During the subsequent integration, the intensity of each reflection in the diffraction image can be determined and an estimate made of its associated uncertainty. Various corrections for variations in background, orientation and detector parameters are applied, any symmetry-equivalent reflections are scaled and multiple observations are merged. Finally, the structure factor amplitudes are estimated from the intensities. As a result of this processing, various statistics can be computed to determine the quality of the available data. Those typically cited include the mosaicity of the crystal, the resolution range, completeness and redundancy of the recorded reflections, as well as the $I/\sigma(I)$ and the R -value for the merged data.

2.5.4.2 *Statistics from scaled and merged data*

As a result of processing the diffraction data, various statistics can be used to assess the quality of the data obtained and to determine information about the crystal structure. The most important statistics are briefly described in the following section.

a) Resolution, completeness, redundancy and mosaicity

The resolution range of the diffraction data greatly affects the quality of any resulting structure and is dependent on the crystal quality, the wavelength of the X-rays used in the experiment, and the data collection parameters (crystal-to-detector distance, exposure etc.). The completeness of the data set may be reduced if a significant amount of data are not recorded, and this may have several possible causes. Many detectors are square, which means that the highest resolution data are recorded in the corners. Some high-resolution reflections may never be detected, because they fall outside the area of the detector.

Another reason for incomplete data is that reciprocal lattice points lying close to the rotation axis will not pass through the Ewald sphere, unless the crystal is also rotated around a second axis. For space groups other than P1 (which has no symmetry), these missing ‘cusp’ data can be minimised by making sure that the symmetry axis of the crystal is not aligned with the rotation axis. In this way, a reflection that lies in the cusp will have a symmetry mate that will pass through the Ewald sphere, and a complete set of data can be collected. Finally, the redundancy refers to the number of observations of each unique reflection.

A further useful statistic is the mosaicity of the crystal. Different parts of the crystal may be slightly misaligned, causing reflections to be diffuse or split. Mosaicity is the recorded rocking width of the reflection with the instrument contribution deconvoluted out, in other words a measure of the spread of each spot in the diffraction

image. It is a global measure that may be caused by various underlying pathologies in the crystal structure. The mosaicity of a crystal appears to be the most commonly used indicator of crystal quality in X-ray crystallography. However, other statistics can tell us about hidden defects, such as twinning, which may not be apparent from the diffraction images.

b) Intensity statistics

The Wilson plot shows the fall of mean intensity with resolution. This takes a predictable form for protein crystals: there is a dip at 5\AA , after which the plot is fairly linear to the diffraction limit. Deviations from an ideal Wilson plot may indicate a problem with the data. Another useful indicator is the cumulative intensity plot, which may be distorted by twinning in the crystal. The moments of the observed normalised amplitudes may also be used to detect the presence of twinning. Intensity statistics compare intensities to the mean intensity, and this is assumed to vary only with resolution, so they may be distorted if there is translational NCS present.

2.5.4.3 *The Matthews coefficient*

The Matthews coefficient (V_M) allows the estimation of the number of molecules in the asymmetric unit. It is given by Equation 4, where V is the unit cell volume, M is the molecular mass of the molecule, N_{ASU} is the number of asymmetric units (depending on the symmetry operators of the space group), and N is the number of molecules in the asymmetric unit. The most probable values of N are those which give Matthews coefficients within the empirically observed range, and data processing programs can give an estimate of the various probabilities by comparing the calculated Matthews coefficient with distributions obtained from published structures. This allows the most likely number of monomers in the asymmetric unit to be estimated.

$$V_M = \frac{V}{M \times N_{ASU} \times N}$$

Equation 4: The Matthews coefficient. V is the unit cell volume, M is the molecular mass, N_{ASU} is the number of asymmetric units (for a given space group) and N is the number of molecules in the asymmetric unit.

2.5.4.4 *The Patterson function*

As previously explained, the structure factor (which has amplitude and phase) and the electron density are related by the Fourier transform (see p. 96). If we ignore the phase information and perform a Fourier transform on the intensity information alone, the result is known as the Patterson function. It is equivalent to the electron density convolved with its inverse, and turns out to be a very useful function.

Peaks in the Patterson function correspond to interatomic vectors weighted by the product of the number of electrons in the atoms in question. The various applications will be discussed further elsewhere, but at this stage the Patterson function can be used to determine the presence of any translational symmetry elements in a crystal. This is because the vectors between equivalent atoms in translationally related molecules will be the same and will add up to give large peaks in the Patterson function. The position of the peaks can give information about the orientation of the symmetry operator.

2.5.4.5 *The self rotation function*

If the Matthews coefficient suggests that there are two or more molecules in the asymmetric unit, the self rotation function can be used to tell if the molecules in the asymmetric unit are related by rotational symmetry. It also allows the orientation of the symmetry axis is to be determined, and this information can be exploited in NCS averaging. Like the Patterson function, the self-rotation function can also be computed using only the unit cell parameters, spacegroup operators, and the scaled and merged intensity data. The self rotation function essentially reveals the rotational symmetry of the Patterson function by comparing two identical copies of the native Patterson map.

Two things affect the appearance of the self rotation function: the radius of integration and the resolution range used in the calculation. The radius of integration excludes points further than a certain distance from the origin of the native Patterson map. A rule of thumb is to start with a radius equal to the diameter of the molecule. The aim is to include as many self-vectors as possible (a larger radius means sharper peaks) while keeping the number of cross-vectors to a minimum, which can obscure any symmetry present in the self-vectors (Doubl  , 2007).

2.5.5 Phasing

One vital piece of information that is required to transform the measured intensities from the diffraction experiment into a map of the electron density within the unit cell is the phase of each reflection. However, this crucial phase information cannot be recorded experimentally. The resulting conundrum is known as the phase problem, and presents a particular challenge to macromolecular crystallographers.

In the case of small molecules, the phases can be estimated by exploiting the assumptions that the electron density must be a positive real number and that the crystal consists of discrete atoms. The family of methods used to do this are known as ‘direct methods’, and while they work well for smaller molecules, their application to much larger and more complex macromolecules is limited. Instead, protein crystallographers must generally obtain phase information by other methods.

Three other principal techniques exist: molecular replacement relies on the availability of a similar structure to the unknown structure; single or multiple isomorphous replacement requires the attachment of atoms with high atomic number to the protein molecules within the crystal; and anomalous dispersion methods depend on the presence of anomalous scatterers in the protein structure itself. The following section will describe the methods explored in this work in more detail.

2.5.5.1 *Direct methods*

Although direct methods are not commonly used to phase protein crystals, they are quite appropriate for molecules of up to ~1000 atoms, and smaller peptides do fall into this category. The basis of direct methods as applied to protein phases is the triplet relation between phase angles, which stems from the fact that structure factors are not independent from each other.

For a reflection H , with indices $(h\ k\ l)$ the structure factor can be expressed in the following way, where Φ_H is the phase of reflection H ; N is the number of atoms in the unit cell; f_j is the scattering factor of atom j ; and x_j, y_j, z_j is the position of atom j in Cartesian coordinates:

Equation 5: Structure factor equation

$$F_H = |F_H| \exp(i\phi_H) = \sum_{j=1}^N f_j \exp 2\pi i(hx_j + ky_j + lz_j)$$

The intensities of reflections are directly related to the structure factor magnitude. It follows that for strong reflections, where F_H is large, then $(hx_j + ky_j + lz_j)$ modulo 1 will be approximately constant for all j . In other words, for H to be a strong reflection, all of the contributing atoms must lie near the corresponding lattice plane. For weak reflections, the atoms are randomly distributed, as shown in Figure 2.10 (Schenk, 2007). For two strong reflections, H and K , the electron density is likely to be found near the intersection of the two corresponding sets of planes. If the $-H$ - K reflection is strong as well, it is likely that the three sets of planes intersect at a common point. This situation is illustrated in Figure 2.11. Taking an arbitrary origin O within the system, the phase angle of each reflection (ϕ) can be calculated as the distance from O to the nearest member of the set of reflecting planes. From this, the triplet relation between phase angles can be derived, which has a probability character (due to the different possible configurations of planes as shown below):

Equation 6: Triplet phase relationship

$$\phi_H + \phi_K + \phi_{-H-K} \approx 0$$

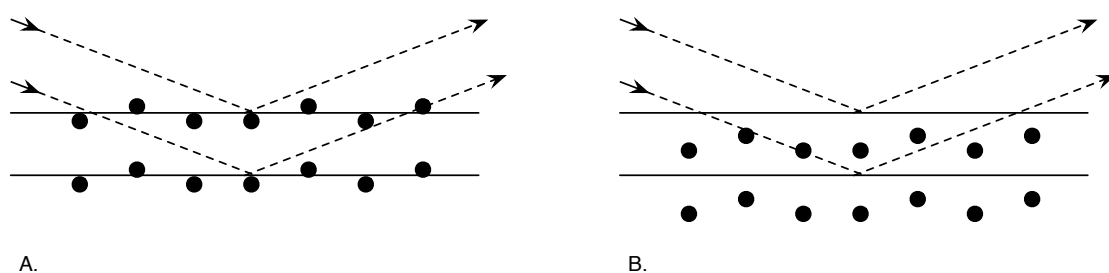


Figure 2.10: Distribution of atoms contributing to strong and weak reflections. A. The reflection $(h k l)$ is strong when the atoms lie near the set of planes $(h k l)$. B. A reflection is weak when the atoms are spread out with respect to the corresponding lattice planes.

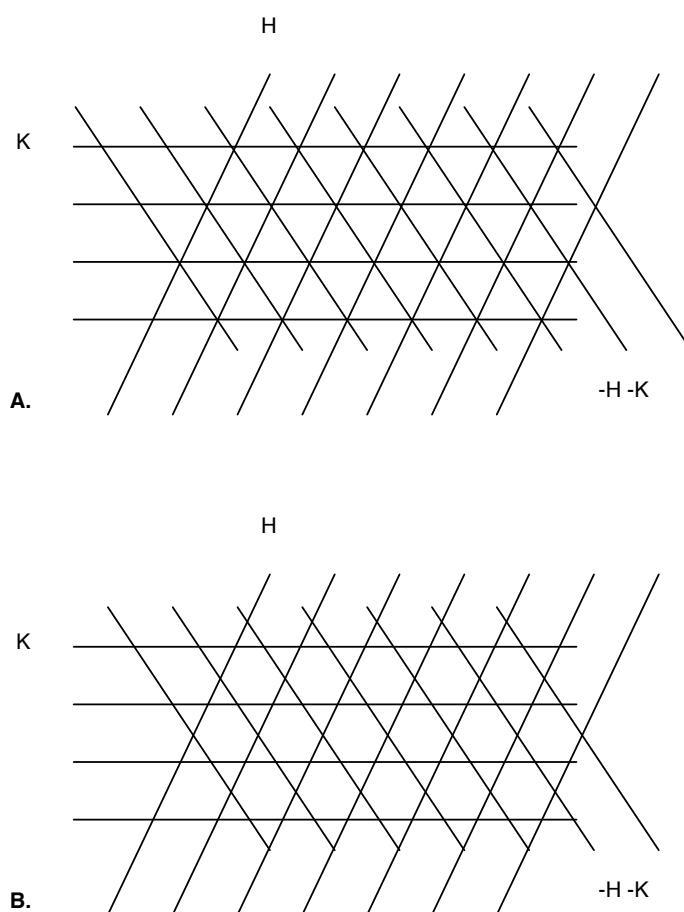


Figure 2.11: Three sets of lattice planes perpendicular to the zone axis. A. When the reflections H , K and $-H -K$ are all strong, it is more likely that the three sets of planes intersect at a common point. B. For weak reflections, the planes are less likely to intersect at a single point.

While the direct application of this triplet phase relationship works well for determining the phases of small compounds, several modifications are required for successful application to protein structures. These will not be covered here, but for further reading see Drenth, 2002. Essentially, the overall procedure for proteins involves generating a small subset of phases (by MR or from the positions of any heavy atom), and using phase relationships to extend and refine the remaining phases. A major prerequisite for the use of direct methods is the availability of atomic resolution data, and the procedure itself is very computationally intensive.

2.5.5.2 *Molecular replacement*

The most common method for solving protein phases is molecular replacement, which is currently used for up to 70% of the structures deposited in the PDB (Evans and McCoy, 2008). In principle, MR involves using a known structure that is sufficiently similar to the unknown structure to provide initial estimates of the phases. A suitable model is rotated and translated through all possible orientations and the predicted diffraction is compared with the experimentally determined intensities. The best fit of the model to the unknown target structure is found where the predicted diffraction and observed diffraction are most similar. The phases from the model are then copied from the model and combined with the experimental amplitudes to calculate an initial electron density map. From there, the phase problem becomes a crystallographic refinement problem. In contrast to other methods, MR offers the advantages of being fast, cheap and highly automated, as well as becoming more applicable as the database of known structures expands.

The first step in an MR procedure is to select and prepare an appropriate model. A good model will have a high similarity to the target structure, which is generally indicated by a high degree of sequence identity, so models are generally selected on the basis of sequence similarity searches. The next step typically involves improving the model by removing parts of the structure that could contribute noise to the search, such as regions of sequence diversity, loop regions and side chains, which can be truncated to common atoms. Depending what is known about the target structure, it is possible to split the search model into individual domains that are used as separate MR models if conformational flexibility is expected. Modification of B factors — which describe the degree to which electron density is spread out and thus indicate the mobility of the atom — can also help to improve models. However, while the success of MR searches depends heavily on having a good enough model, it is impossible to quantify ‘good enough’ from the outset of an MR procedure.

Since a molecule's orientation is defined by six parameters — three rotation angles and three translations within the unit cell — a full six-dimensional search can take an enormous amount of time and computational power. For this reason, the actual MR search is performed in two steps: a rotation search is carried out first, and the best solutions from this are then tested in a translation search. The MR rotation search can be conducted in one of two general ways, using either traditional Patterson methods or a 'maximum likelihood' statistical approach.

The Patterson function is essentially the Fourier transform of the observed intensities. A map of the Patterson function corresponds to a map of the interatomic vectors with the weights of the vectors being proportional to the atomic scattering. For MR, the Patterson maps of the known and unknown structures are calculated and the model Patterson map is rotated with respect to the Patterson map calculated from the observed data to find the best correlation. However, since the Patterson map contains both intramolecular self-vectors and intermolecular cross-vectors between symmetry-related molecules, the Patterson map for the MR search model is calculated in an extremely large P1 cell. This means that only the self-vectors will be present, thus simplifying the search, and the search is performed within a defined radius from the origin of the Patterson function.

In contrast to Patterson-based methods, maximum likelihood methods involve probability functions in reciprocal space. Essentially, this approach asks the question: what is the probability of obtaining the observed structure amplitudes for any proposed orientation (rotation and translation) of the search model? The likelihood for the model is calculated on a grid of angles or positions and the most likely orientation of the search model — given the observed data — is selected as the MR solution. The mathematics behind this approach to MR will not be covered in detail here, but for a comprehensive overview see McCoy, 2004. Maximum-likelihood MR explicitly models errors both in the experimental data and the search model, which is one reason why this approach is generally more robust than Patterson-based MR.

In both cases, the user is able to fine-tune the parameters of the search, in addition to manipulation of the model and the input data, which can dramatically affect the effectiveness of the search. Since the bulk of structural information is contained within the phases, the model used for the MR search contributes considerably to the final solution. In order to determine whether the solution is correct, maps phased from the solution can be inspected for new, plausible features that were not present in the model structure. Alternatively, regions of the model can be omitted to see if they reappear. Overall, MR offers a relatively rapid and powerful method of obtaining protein phases, provided a suitable model of the target structure is available.

2.5.5.3 *Isomorphous replacement*

Another method of obtaining phases for protein structures is through isomorphous replacement, which comes in two flavours: multiple isomorphous replacement (MIR) and single isomorphous replacement with anomalous signal (SIRAS). These techniques rely on perturbing the structure factors in a way that allows the phases to be deduced.

All the electrons in an atom scatter roughly in phase and the scattering intensity is proportional to the square of the atomic number. Most of the atoms in a protein are relatively light elements (carbon, oxygen, nitrogen, hydrogen) and the contributions from these atoms will tend to cancel out, since each atom will scatter with a different phase angle. Introducing heavy atoms (i.e. elements possessing many electrons) into the crystal structure significantly alters the resulting scattered intensities, as shown in Figure 2.12.

Only a few heavy atoms are required to produce measurable differences in the scattered intensities, and it is assumed that the addition of heavy atoms does not perturb the protein structure: in other words, a crystal containing only protein and a crystal of the same protein containing heavy atoms will be isomorphous. This is typically accomplished by soaking protein crystals in solutions of heavy metal salts.

The difference between the scattered intensities from a native crystal and a heavy-atom derivative largely reflects the scattering contribution of the heavy atoms. The positions of a few heavy atoms can easily be determined, for example by computing a Patterson map or by direct methods. Once the positions of the heavy atoms within the crystal are known, it is possible to deduce the unknown phases of the native protein structure.

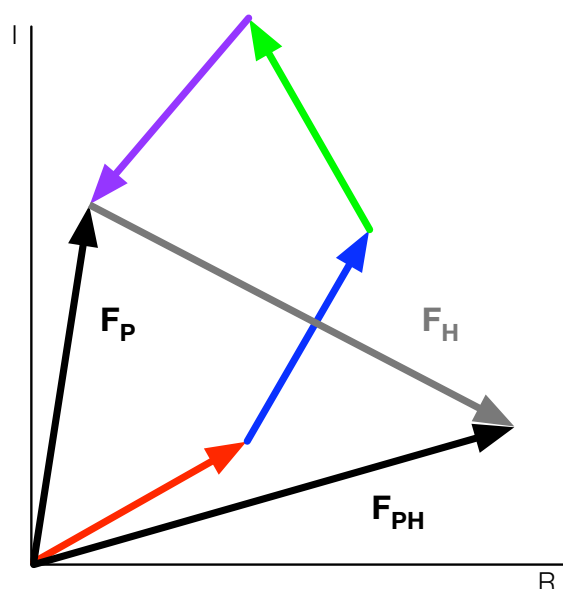


Figure 2.12: The effect of introducing heavy atoms on the resulting structure factor. Contributions from four individual light atoms are shown in colour. Each atom scatters with a different phase angle, and the structure factor is given by F_P . The presence of heavy atoms within the crystal structure causes a large change to the resulting structure factor, F_{PH} . The contribution of the heavy atoms is given by F_H .

However, there are two possible solutions for the phase. This can be explained by viewing the structure factors as vectors. The magnitude and direction of the vector F_H is known from the calculated phases and amplitudes of the heavy atom positions. Furthermore, the magnitude of vectors F_P and F_{PH} are known from the observed intensities of the native and derivative diffraction data. This means that the problem of determining the unknown phase of F_P reduces to the problem of constructing a triangle connecting the vectors, given the following relationship:

$$F_{PH} = F_P + F_H$$

Equation 7: Relationship between structure factors of native and derivative crystals. The structure factor for the heavy atom derivative, F_{PH} , is equal to the sum of the structure factors from the native crystal, F_P , and the heavy atom contribution, F_H .

There are two possible ways in which to construct such a triangle, which give two possible phase angles for F_P and thus two possible sets of phases for the protein structure. More information is therefore required in order to assign the phase unambiguously. This information can be derived from another isomorphous derivative (as in MIR) or through collecting anomalous scattering data (as in SIRAS).

Anomalous dispersion of X-rays by heavy atoms can cause a breakdown in Friedel's law (see Equation 3, p. 96). This can be exploited by SIRAS to remove the need for a second derivative, and by MAD to obtain analogous information to that gained from MIR. The strength of the anomalous scattering effect depends on the wavelength of the X-rays being similar to the natural frequency for the heavy atom. The strong scatterers may be intrinsic to the protein structure or they may be introduced, for example by replacing methionine residues with selenomethionine. By collecting data at several wavelengths near the absorption edge of the element in the crystal, phase information can be obtained. However, since anomalous scattering methods were not used in this work, they will not be further discussed.

2.5.6 Density modification

Once phases have been obtained by one of the methods outlined above, it is often necessary to improve them. Depending on the quality of the diffraction data, in combination with the quality of the MR model or the heavy atom positions, there may be large errors in the resulting electron density maps. However, there are certain features that are characteristic of a good electron density map. The techniques collectively known as density modification seek to improve the calculated phases by modifying the map to look like an ideal electron density map.

Protein crystals have a high solvent content, with typical crystals containing around 50 % solvent (McPherson, 2004). This means that approximately half of the unit cell volume is likely to be disordered solvent, which should have flat, featureless electron density. Solvent flattening is one procedure for improving the phases, which works by adjusting them to produce featureless solvent regions.

Another means of improving phases assumes that when the same molecule appears in different regions of the electron density map (for example due to non-crystallographic symmetry), the electron density of each copy is the same. By merging the information from each copy, any random phase errors will be reduced in a process called averaging.

Finally, we know that proteins are made up of similar atom types and therefore different proteins will yield similar distributions of electron density. Histogram matching essentially matches the distribution of high and low densities in the protein region with the expected distribution. Any differences are most likely due to phase errors and so phases can be improved by altering the calculated electron density distribution to more closely match the expected values.

2.5.7 Model building and refinement

Once phase information has been obtained and an electron density map has been calculated for the data, the next step is to build a model of the protein structure. If MR has been used to solve the phase problem, then a relatively accurate model of the structure has already been placed within the electron density map. However, the map will be heavily biased towards the model used in the MR procedure, since the phases calculated from the model contain more information than the observed intensities. In this case, any missing or incorrect regions of the model must be fitted and the model must be refined against the observed data in order to remove model bias.

If the phase information was gained by other methods, it is first necessary to build a model of the protein. This is often done using automated model building and refinement programs. These exploit the information we know about protein structures (bond lengths, angles, the amino acid sequence) to build an atomic model that best fits the observed electron density given these parameters. A preliminary structure is built in interpretable regions of electron density, and through an iterative procedure of model building and refinement (possibly combined with further rounds of density modification) the model can be improved until it is complete or suitable for manual inspection.

Refinement is the process of improving the model by shifting the positions, occupancies and temperature factors of atoms within it so that the model explains the observed diffraction data better. Until relatively recently, refining a model proceeded by minimising the sum of squares of deviations between the observed and calculated intensities. However, maximum likelihood methods are now commonly used instead of least-squares minimisation. These rely on the same principles as the maximum likelihood methods used in MR: asking how probable the observed intensities are from a given model. By altering the model to make the observations more probable, the

agreement between the model and the data is improved, resulting in a better model. The measure of agreement is expressed by the R-factor, which is given by Equation 8. The diffraction intensities are proportional to the squared structure factor magnitudes.

$$R = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|}$$

Equation 8: The crystallographic R-factor. The sum of the structure factor amplitudes extends over all measured reflections and their calculated counterparts.

Despite using stereochemical restraints during refinement, it is possible to over-fit or misfit the model. This can give acceptable R-factor values even with an incorrect model. So a modified R-factor, known as the free R value (R_{free}), can be used to follow a refinement procedure. The free R value expresses the agreement between observed and calculated structure factor amplitudes for a ‘test’ set of reflections omitted during the modelling and refinement process (Brunger, 1992). There is a high correlation between R_{free} and the accuracy of the model phases, and the free R value is generally believed to be a less biased measure of refinement than the R-factor. Generally, both the R-factor and R_{free} are used to guide the refinement process, since any discrepancy between the two could be a useful indicator in the search for accurate and correct models of protein structures.

Refinement is the last step in solving a protein structure by X-ray crystallography, and it is only really at this stage that all of the assumptions made throughout the modelling process can be shown to be correct. When a model is refined against a random set of reflections, we can expect an R-factor of around 0.6, which indicates that the model does not accurately predict the data. If the R-factor stays high, this may indicate an incorrectly assigned space group or problems with the data or the model. A correct protein structure will typically have an R-factor of around 0.2 (for a well refined model at a resolution of 2.5 Å).

Chapter 3

Materials and methods

3.1 Purification and characterisation of C1

3.1.1 Materials

Frozen human plasma for the preparation of C1 was obtained from National Blood Service (Southampton, UK). Plasma was split into 50ml aliquots, which were stored at -20°C. Chemicals were obtained from Sigma (Sigma-Aldrich Company Ltd., UK) unless otherwise stated. Crystallisation reagents were purchased from either Hampton Research, USA, or Molecular Dimensions, UK.

3.1.2 Bioinformatics

The sequence of an assembled molecule of human C1 was constructed using sequences retrieved from the UniProtKB/Swiss-Prot database (2010). The assembled C1 sequence was composed of 6 copies each of the C1q A, B and C sequences (accession numbers P02745, P02746 and P02747), as well as 2 copies each of the C1r (P00736) and C1s (P09871) sequences. The assembled C1 sequence was analysed with Protparam (Wilkins et al., 1999). Sequence based metal binding site prediction was carried out using sequences of the individual C1 subcomponents and MetalDetector 2.0 (Lippi et al., 2008).

3.1.3 Preparation of serum from human plasma

An aliquot of plasma was thawed and calcified by the addition of 50mM calcium chloride. The sample was left to clot at room temperature for one hour and then overnight at 4°C. The clot was removed by centrifugation at $6000 \times g$ for 1 hour leaving the supernatant serum.

3.1.4 Precipitation of C1 from human serum

C1 was precipitated from human plasma or serum according to the method of Neoh et al., 1984 to prepare C1-deficient serum. PEG 6000 was slowly added to volume of plasma or serum to a final concentration of 3.5% under constant slow stirring. The precipitation was carried out on ice, and allowed to proceed for 30 minutes. The precipitate was recovered by centrifugation at $6000 \times g$ for 30 minutes.

3.1.5 Purification of C1 by SEC

The precipitate containing C1 obtained by above procedure was re-suspended in 1/10–1/50 the original volume of buffer and clarified by centrifugation at $6000 \times g$ for 10 minutes. Buffers used in the various experiments detailed in Chapter 4 are shown in Table 5. The supernatant was removed and filtered through a 0.2 μm filter before being loaded onto a Superose 6 10/300 GL Tricorn™ high performance column (GE Healthcare, UK). All chromatography steps were carried out using the ÄKTA FPLC™ system (GE Healthcare, UK) and chilled buffers. Isocratic separation was performed at

0.5ml/min in the same buffer used to re-suspend the sample and 1ml peak fractions were collected. The peak containing C1 was determined by analysing the collected fractions by SDS-PAGE, as described on page 126.

Buffer (10mM, pH 7.4)	NaCl (mM)
Tris	–
	150
	500
Phosphate	150
HEPES	150
MES	150
MOPS	150

Table 4: Buffers used to re-suspend precipitated complement component C1. The buffer used to re-suspend the PEG-precipitated protein was matched to the buffer used in the subsequent purification by size exclusion chromatography.

3.1.6 Further purification by IEX

Fractions containing C1 from the SEC purification were pooled and exchanged into 10mM Tris, pH8, 150mM NaCl. A Mono QTM 5/50 GL column (GE Healthcare, UK) was equilibrated with the same buffer and the sample was applied. The column was washed until a stable baseline was achieved and a gradient elution was performed using 10mM Tris, pH8, 500mM NaCl (buffer B) to elute the bound proteins. Various gradients used in this work are given in Table 5. Peaks eluting at 18% and 23% B were collected in 0.5ml fractions.

Length (CV)	Gradient
20	0–100 % B
20	0–100 % B; held on peaks

Table 5: Gradients used in the purification of complement component C1 by IEX. Using a stepped gradient gives a cleaner separation between peaks in the chromatogram. This process can be automated so that the gradient is held when a peak is detected.

3.1.7 Electrophoresis

3.1.7.1 Native agarose gel electrophoresis

Samples of oxidised and reduced C1, as well as marker proteins thyroglobulin (669kDa, pI 4.6), ferritin (440kDa, pI 4.4) and catalase (232kDa, pI 5.6) were prepared as follows. A sample of C1 purified by SEC (see page 124) was reduced by incubating with 0.1 vol. BME for 10 min at room temperature. Stock solutions of the marker proteins were prepared in 10mM Tris, pH 7.4, 150mM NaCl. Samples were prepared to contain comparable amounts of protein and mixed 1:1 with loading buffer (200mM Tris, pH 8.5, 20% glycerol, 0.010% bromophenol blue). Separation was performed on a 1% agarose gel, equilibrated with 100mM Tris, pH 8.5, at 20mA for 1.5 hours and gels were stained using Coomassie as described on page 127.

3.1.7.2 SDS-PAGE

Separations were carried out according to the standard Laemmli protocol on 12% polyacrylamide gels.

3.1.7.3 *Staining*

Gels were stained using the colloidal Coomassie Brilliant Blue staining method described by Candiano et al., 2004. Polyacrylamide gels were stained for several hours using a colloidal suspension of 0.12% Coomassie Blue G-250 in 20% methanol, 10% phosphoric acid and 10% ammonium sulphate. No de-staining step was necessary with this protocol and gels stained by this method are fully compatible with mass spectrometry.

After staining with Coomassie, gels could subsequently be stained with silver according to the method outlined by Bollag et al., 1996. Gels stained by this method are not compatible with mass spectrometry but the method was chosen since all of the required reagents were readily available.

3.1.8 Mass spectrometry

Proteins were sent for analysis by peptide mass fingerprinting at the Sussex Proteomics Centre (UK). Samples were provided either as solution samples in low-salt aqueous buffers (e.g. 10mM Tris, pH 7.4, 150mM NaCl) or as gel slices for in-gel digestion at the facility. For in-gel digestion of individual protein bands, proteins electrophoresed on polyacrylamide gels according to the method outlined on page 126 and stained using Coomassie Brilliant Blue. Gel slices were prepared according to the two methods described below.

In the first method, the stained gel was placed on a glass plate and the desired bands were excised using a scalpel, making sure to clean the scalpel between each

excision to prevent contamination. Once excised, each gel slice was cut into several smaller pieces, again taking care to prevent contamination. The gel slices were placed into clean, sealed containers for transport and samples were processed the same day.

The second method involved the same procedure described above, except that the excision of the gel pieces was performed under a laminar flow hood. In both methods all glassware and tools were thoroughly cleaned using detergent, followed by a wipe with methanol and lint-free tissue. Reagents were freshly prepared and gloves were worn throughout the electrophoresis run and the remaining sample preparation. Care was taken to minimise exposure of the gel to dust during staining and handling.

All samples were analysed using Sequest (Thermo Fisher Scientific, USA) and X! Tandem (The GPM, thegpm.org). Sequest was set up to search the ipi.HUMAN.v3.72 database, assuming the digestion enzyme trypsin. X! Tandem was set up to search a subset of the ipi.HUMAN.v3.72 database also assuming trypsin. Scaffold (Proteome Software Inc., USA) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability as specified by the Peptide Prophet algorithm (Keller et al., 2002). Protein identifications were accepted if they could be established at greater than 50.0% probability and contained at least two identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii et al., 2003).

3.1.9 Dynamic light scattering

All DLS measurements were taken on a Zetasizer Nano S (Malvern Instruments, UK) calibrated using 60nm latex beads according to the manufacturer's instructions. Samples eluting directly from the chromatography column were not filtered before being analysed, since purification buffers were all pre-filtered through a 0.2 μm filter, and the FPLC system and columns also contain several further filters. In all other cases, samples were filtered through a 0.2 μm filter prior to analysis. All attempts were made to protect the samples from contamination by dust and other particles.

Experiments to determine the effect of ionic strength, EDTA, calcium etc. on purified C1 were carried out as follows. The various test solutions were prepared and a sample of purified C1 was added to the desired solution so that the final concentration of C1 was 100 $\mu\text{g}/\text{ml}$, unless otherwise stated. The melting point of purified C1 was determined using the integrated melting point determination function of the software accompanying the instrument. Temperature was ramped up in either 1°C or 5°C intervals and three measurements were taken at each temperature step, which were subsequently averaged.

3.1.10 TEM

Preparation of C1 samples for analysis by transmission electron microscopy was carried out according to a modified version of the method of Strang et al., 1982. C1 was purified in 10mM MOPS, pH7, 150mM NaCl, and diluted to 100 $\mu\text{g}/\text{ml}$. EDAC was prepared at 10mg/ml in distilled water and mixed with purified C1 in a 1:10 ratio of EDAC to protein. The reaction was allowed to proceed at room temperature for 1 hour and terminated by exchanging the sample into 10mM Tris, pH 7, 150mM NaCl.

A small volume of cross-linked protein sample was dispensed onto a carbon coated TEM support grid and negatively stained with 1% aqueous uranyl acetate. A control sample containing only 10mM Tris, pH 7, 150mM NaCl and no protein was prepared in the same fashion. The samples were analysed on a Hitachi-7100 Transmission Electron Microscope (TEM) with an axially-mounted Gatan Ultrascan 1000 CCD camera.

3.1.11 Crystallisation

Appropriate protein concentration ranges for crystallisation were determined using the pre-crystallisation test (Hampton Research, USA) according to the manufacturer's instructions. Crystallisation experiments were performed by the hanging drop method, using 1µl of purified C1 and a 0.5 ml reservoir. The experiments were set up at room temperature and transferred to 4°C.

3.2 Investigating the ligand binding properties of SAP

3.2.1 Materials

Frozen human plasma for the preparation of SAP was obtained from National Blood Service (Southampton, UK). The plasma was thawed and split into 50ml aliquots, which were stored at -20°C. Chemicals were obtained from Sigma (Sigma-Aldrich Company Ltd., UK) unless otherwise stated.

3.2.2 Purification of SAP

The purification scheme used to prepare SAP was adapted from that of De Beer and Pepys, 1982. Human plasma was incubated overnight at 4°C under constant slow stirring with Sepharose 4B equilibrated with 10mM Tris, pH7.4, 150mM NaCl. The slurry was transferred to a clean XK 50/60 column (GE Healthcare, UK) and washed with equilibration buffer until a stable baseline was achieved. Bound SAP was eluted with by flushing the column with 10mM Tris, pH7.4, 150mM NaCl, 50mM EDTA. Peak fractions were collected and pooled.

The crude SAP preparation obtained by the affinity step described above was further purified by anion exchange chromatography. A Mono Q™ 5/50 GL column (GE Healthcare, UK) was equilibrated with 10mM Tris, pH7.4, 150mM NaCl and the pooled fractions from the previous step were loaded on to the column. Any unbound proteins were washed from the column using equilibration buffer and SAP was eluted using a 20 CV linear gradient of 10mM Tris, pH7.4, 1M NaCl. Peak fractions were collected and analysed by SDS-PAGE on 12% gels using the standard Laemmli protocol.

3.2.3 Aggregation assays

DLS was performed using the Zetasizer Nano S (Malvern Instruments, UK) calibrated using 60nm latex beads according to the manufacturer's instructions. All samples were filtered through a 0.2 μm filter prior to analysis.

A sample of purified SAP was buffer exchanged into 10mM Tris pH8, 140mM NaCl and diluted to a final concentration of 50 $\mu\text{g/ml}$ SAP. Calcium-dependent aggregation of SAP was induced by the addition of 1mM calcium, pH7.4. The sample was immediately placed in the DLS instrument and measurements were taken at 5 minute intervals for 1 hour.

After adding 1mM calcium, a further sample of SAP was incubated at room temperature for one hour to allow aggregates to form. PE was then added to a final concentration of 1mM and measurements were made as above. Finally, 1mM PE was added first to a solution of SAP, followed by the same concentration of calcium before measurements were taken.

3.3 Methods: ZiCo

3.3.1 Materials

ZiCo95 was obtained as a 1.1 mM stock solution in water from the Woolfson laboratory (Bristol, UK). ZiCo1 was obtained lyophilised and frozen at -20°C from Eleonora Cerasoli (Sussex, UK) and Beth Bromley (Bristol, UK). Stock solutions of 5–10 mg/ml were prepared by dissolving the peptide in water, and stored at 4°C. Chemicals were obtained from Sigma (Sigma-Aldrich Company Ltd., UK) unless otherwise stated.

3.3.2 ZiCo95

3.3.2.1 Crystallisation

The Pre-Crystallisation Test (Hampton Research, USA) was set up according to the manufacturer's instructions using 1 µl drops of all the various combinations of peptide concentration, peptide:zinc ratios and sources of zinc shown in Table 6.

ZiCo95 (mM)	Peptide:zinc	Zinc source
0.1	1:0	ZnCl ₂
0.5	1:1	ZnSO ₄
1	1:5	
	1:10	

Table 6: Variables used in the PCT of ZiCo95. All combinations of the factors shown were tested to determine an optimal starting point for crystallisation trials of the peptide.

Crystallisation trials of ZiCo95 were set up using the Stura Footprint screen (Molecular Dimensions) and the hanging drop method using 1 μ l drops over a 0.5ml reservoir. Trays were set up at room temperature and then either transferred to 4°C or placed in a polystyrene box at room temperature, to minimise thermal fluctuations. The reducing agent BME was included as a parameter in the crystallisation trials at a concentration of 1 mM in the crystallisation drop. In addition, experiments were set up using different peptide concentrations (0.3, 0.5 and 0.8mM), crystallisation buffers (150mM Tris or 150mM phosphate) and temperatures (4°C and room temperature).

Promising conditions from the initial screens were identified where an increase in precipitant concentration resulted in a transition from a clear drop to precipitate. These promising conditions were then expanded into further screens by varying the precipitant concentration within the identified range and the pH by ± 1 unit around the identified condition.

3.3.2.2 SEC

A Superdex Peptide 10/300 GL column (GE Healthcare, UK) was calibrated according to the manufacturer's instructions. Separations were performed in 50mM Tris, pH 7.5, 50mM sodium chloride, using 20nmol of peptide per run. For reduced samples, 10mM BME was included in the sample buffer.

3.3.3 *ZiCo1*

3.3.3.1 *Crystallisation*

The Pre-Crystallisation Test (Hampton Research, USA) was carried out as for ZiCo95 (see page 133). Crystallisation trials were set up using the Stura Footprint Screen (Molecular Dimensions, UK). Again the hanging drop method was employed, using 1 μ l drops over a 0.5ml reservoir. Trays were set up at room temperature and then either transferred to 4°C or placed in a polystyrene box at room temperature, to minimise thermal fluctuations.

Optimisation of crystallisation conditions involved expanding initial hits into further 4×6 screens in which the precipitant concentration and pH were incrementally varied (see Table 33, p. 261). Growth of large, single crystals using sodium citrate as a precipitant was encouraged by streak seeding into drops containing slightly lower precipitate concentrations than those required for spontaneous nucleation (< 0.75 M). This was done by lightly scratching the surface of a crystal with an acupuncture needle and then touching the needle to a freshly prepared drop. The drop was placed over a reservoir and incubated until crystals were observed.

3.3.3.2 *CD*

Circular dichroism spectroscopy was performed on a JASCO J-715 CD spectropolarimeter (JASCO, UK). A sample of 100 μ M ZiCo in 50mM sodium phosphate, pH 7.5, 50mM sodium chloride was analysed with or without equimolar zinc chloride present in the sample. Analysis was carried out according to the manufacturer's instructions.

3.3.4 Crystallography of ZiCo1

Data were collected using a Rigaku RU-H3RHB rotating anode X-ray generator equipped with Osmic Max-Flux optics and an RAXIS-IV++ image plate area detector. Cryocrystallography was carried out at 100K using the MSC X-Stream cryo system. Subsequent computational steps were carried out using various programs from the CCP4 software suite (Collaborative Computational Project, 1994), as described in the relevant sections. Models and maps were viewed and manipulated using the molecular graphics program Coot (Emsley and Cowtan, 2004).

3.3.4.1 *Cryoprotectant screening*

Suitable conditions for cryoprotection of the crystals were identified as follows. A sample of the mother liquor surrounding the crystals in the crystallisation drop was flash frozen and two test shots were made (oscillation angle: 1°; exposure: 1 min; 90° apart). The presence of ice rings in the diffraction pattern indicates the need for additional cryoprotectants.

Various cryoprotectants from the CryoPro kit (Hampton Research, USA) were tested, including glycerol, MPD and low molecular weight PEGs. The cryoprotectant was added to well solution in 10% concentration steps until adequate cryoprotection was achieved, confirmed by the absence of ice rings in the diffraction pattern. Crystals were either immersed briefly in the final cryoprotectant solution, or transferred through a series of solutions with increasing cryoprotectant concentrations.

3.3.4.2 *Cryocrystallography*

A single crystal grown in 0.75M sodium citrate, 10mM borate, pH8.5 was collected and briefly immersed in a solution of crystallisation buffer containing 33% glycerol as a cryoprotectant. The crystal was then flash frozen at 100K directly in the aligned cryostream. A total of 180 images with an oscillation angle of 1° and an exposure of 1 min were collected.

3.3.4.3 *Room temperature crystallography*

Another crystal grown in 0.75M sodium citrate, 10mM borate, pH8.5 was mounted at room temperature according to the method of Mac Sweeney and D'Arcy, 2003. Briefly, the method involves mounting a crystal in a loop as for cryocrystallography and then placing the loop inside a sealed glass capillary tube. Two orthogonal 10° wedges of data were collected at room temperature with an oscillation angle of 1° and exposure of 10 min.

3.3.4.4 *MR models*

a) Zinc finger models

The starting point for zinc finger models for molecular replacement was the structure 1ZNF, a 27-residue classical zinc finger structure determined by NMR. A single model from the ensemble contained within the PDB file was selected and all residues were truncated to alanine. Disordered residues from the N and C termini were also removed, using cartoon models to determine ordered regions of secondary structure.

b) Coiled coil models

Various trimeric coiled coil structures were identified by a manual trawl through the PDB and the crystal structure 1GCM was chosen as the basis for constructing coiled coil molecular replacement models. 1GCM is a parallel, trimeric coiled coil variant of the GCN4 leucine-zipper dimerisation domain. It was selected as a suitable model for the proposed ZiCo1 structure based on sequence similarities. Sequence alignment was carried out using CLUSTAL 2.0.11.

An initial series of models was prepared by selecting a single helix from the 1GCM structure and truncating all residues to alanine. Chains of various lengths from 9 to 27 residues were generated by removing a number of residues from the N terminus. Key residue models were constructed from a single helix of the 1GCM structure as follows: conserved residues were left in place, conserved and semi-conserved substitutions were truncated to the last common atom and all other residues were truncated to alanine. The Phe10 model was constructed from a 7-residue polyalanine helix by mutating one of the central residues to phenylalanine using Coot (Emsley and Cowtan, 2004). Trimeric coiled coil models were generated by truncating the side chains of all three chains of 1GCM to alanine. Models between 15 and 27 residues in length were created by sequentially removing three residues from the N terminus.

3.3.4.5 Heavy atom soaks

The following compounds were selected from the Heavy Atom Screens (Hampton Research, USA), based on their solubility and binding characteristics at the relatively high pH (8.5) of the crystallisation solution: K_2PtCl_4 , $\text{K}_2\text{Pt}(\text{NO}_2)_4$, $\text{K}_2\text{Pt}(\text{CN})_4$, $\text{KAu}(\text{CN})_4$, CdCl_2 and NaIrCl_6 . Stock solutions (10mM) of the compounds were prepared in 1M sodium citrate, 10mM sodium borate, pH8.5. The concentration of precipitant in the heavy atom solution was increased over that in the crystallisation

solution, to compensate for the lack of protein. A control experiment showed that ZiCo1 crystals grown using 0.75 M sodium citrate remained visibly intact after soaking for one hour in the 1 M solution. A small drop of either 1 mM or 10 mM heavy atom solution was placed on a cover slip and a single ZiCo1 crystal was placed in each drop. Crystals were left to soak for varying lengths of time from a few seconds to several hours over a reservoir of well solution and observed at regular intervals.B

Chapter 4

Purification and characterisation of C1

4.1 Introduction

As the initiator of the classical pathway of complement, the first line of defence against invading pathogens or endogenous threats, C1 is of interest from both a purely scientific as well as a medical point of view. The experiments described in the following chapter were designed to examine various physiologically relevant conditions on native human C1 with a view to ultimately obtaining a structure of the entire complex by X-ray crystallography in order to further our understanding of this important protein. As described in Chapter 1, the approach taken here made use of a method of obtaining functionally pure C1 that has not been fully explored in the published literature. So the initial starting point of these investigations was the search for a purification scheme that would allow the native C1 complex to be isolated from human blood plasma in a form suitable for crystallisation. The following sections detail the experiments that were performed on complement component C1 in line with the aims set out for this project in section 1.8.

4.2 Bioinformatic analysis

In order to obtain useful information about the C1 complex before beginning to design a purification strategy, the protein sequence was analysed using bioinformatics tools available on the web as described in section 3.1.2, p. 123. The calculated values for various properties of the assembled C1 complex were useful in the absence of any experimental data and where no other information could be obtained from the literature.

```

Number of amino acids: 7232
Molecular weight: 782835.3
Theoretical pI: 7.25

Extinction coefficients:

Conditions: 6.0 M guanidium hydrochloride
           0.02 M phosphate buffer
           pH 6.5

Extinction coefficients are in units of M-1 cm-1, at 280 nm.

Ext. coefficient    851865
Abs 0.1% (=1 g/l)   1.088, assuming ALL Cys residues appear as half cystines

Ext. coefficient    840240
Abs 0.1% (=1 g/l)   1.073, assuming NO Cys residues appear as half cystines

```

Figure 4.1: Bioinformatic analysis of the complement component C1 sequence using Protparam. The values were calculated using a sequence assembled using the various subcomponent sequences. C1 is predicted to have a molecular weight of around 780kDa and a pI of 7.25. The extinction coefficient of C1 can be assumed to be around 1mg/ml.

Further bioinformatic analysis is complicated by the fact that C1 is a protein complex. When assembled, some parts of the structure may be masked by interactions with other parts of the molecule, while additional features may be formed by the interaction of the

various subunits. The flexibility of the structure (for example the hinge regions of the collagen helix and the modular catalytic core) means that fold prediction algorithms are unlikely to return meaningful results. Using the assembled sequence is only acceptable for the basic primary sequence analysis presented above, since the molecule is not composed of a single contiguous sequence. Sequence based approaches to metal binding site prediction did not return any predicted metal binding sites, while structure based approaches require experimentally determined structures. While predictions of metal binding sites could be performed using the available high-resolution structures of fragments of the complex, the results are unlikely to be meaningful since the fragments themselves do not reflect the true nature of the molecule.

4.3 Purification of native complement component C1

In the following section, the development and optimisation of a novel C1 purification strategy will be described. Although numerous purification strategies have been reported in the literature, the reasons for developing this particular purification procedure will be explained in the context of the aims set out for this project. A simple, two-step purification based on PEG precipitation of human serum proteins followed by size exclusion chromatography had already yielded promising results (Darren Thompson, personal communication), and this formed the basis of the purification scheme developed here.

The approach to purifying C1 outlined below is different from previously published methods, which are largely affinity-based. Previously reported methods of isolating C1 from human serum largely rely on its ability to bind IgG (Arlaud et al., 1980; Medicus and Chapuis, 1980), while more recent studies use recombinant technology to produce modular fragments of the molecule for investigation using high-resolution structural biology techniques (Gregory et al., 2003; Budayova-Spano et al., 2002). The aim of this project is to isolate the entire C1 complex in its native state and in a form suitable for crystallographic studies. This novel method is non-disruptive, versatile, and does not involve a ligand-binding step, which makes it suitable for fulfilling the goals set out above.

4.3.1 PEG precipitation

The starting point for the C1 purification scheme was the preparation of a crude extract containing high-molecular-weight plasma components — including C1 — based on the method of Neoh et al., 1984. This was achieved by slowly adding PEG 6000, under constant slow stirring, to a volume of human plasma to obtain a final concentration of 3.5%. The precipitation was carried out on ice and allowed to proceed for 30 minutes. The precipitate was recovered by centrifugation. Storing the PEG precipitated proteins for extended periods of time at 4°C gave comparable yields of protein to those obtained from freshly prepared precipitate. For practical purposes, 50 ml of plasma was used in a single preparation.

4.3.2 Purification by size exclusion chromatography

The crude extract obtained by the PEG precipitation procedure described above is enriched in C1 but also contains a number of other high-molecular-weight contaminants, which can be separated out by further purification (see Figure 4.2). Size exclusion chromatography was chosen as a suitable step to further fractionate the extract, since this gentle method minimises disruptive forces on the C1 complex and can easily be fine-tuned to produce an optimal purification. Furthermore, the method does not involve binding of the complex to any substrate, since size exclusion media are biologically inert. The following section describes the stages that were carried out in the optimisation of this method.

4.3.2.1 *Choice of size exclusion matrix*

There are a variety of matrices available to choose from, with different properties that must be selected to best fit the requirements of the experiment. Two different gel filtration mediums were investigated for their suitability in the purification of C1. The first, Superdex 200, contains dextran chains covalently attached to a highly cross-linked agarose gel and separates proteins between 10–600kDa. The second medium chosen for the C1 purification was Superose 6, a purely agarose-based medium that offers separation in the range 5–5000kDa. The two separations are compared in Figure 4.2. While the majority of contaminants could be removed using either one, the greater separation range of Superose 6 mean that better separations could be achieved using this medium, and so all further work was carried out on a Superose 6 10/300 GL Tricorn™ high performance column (GE Healthcare, UK). The column was calibrated as described in Appendix I, section ii.

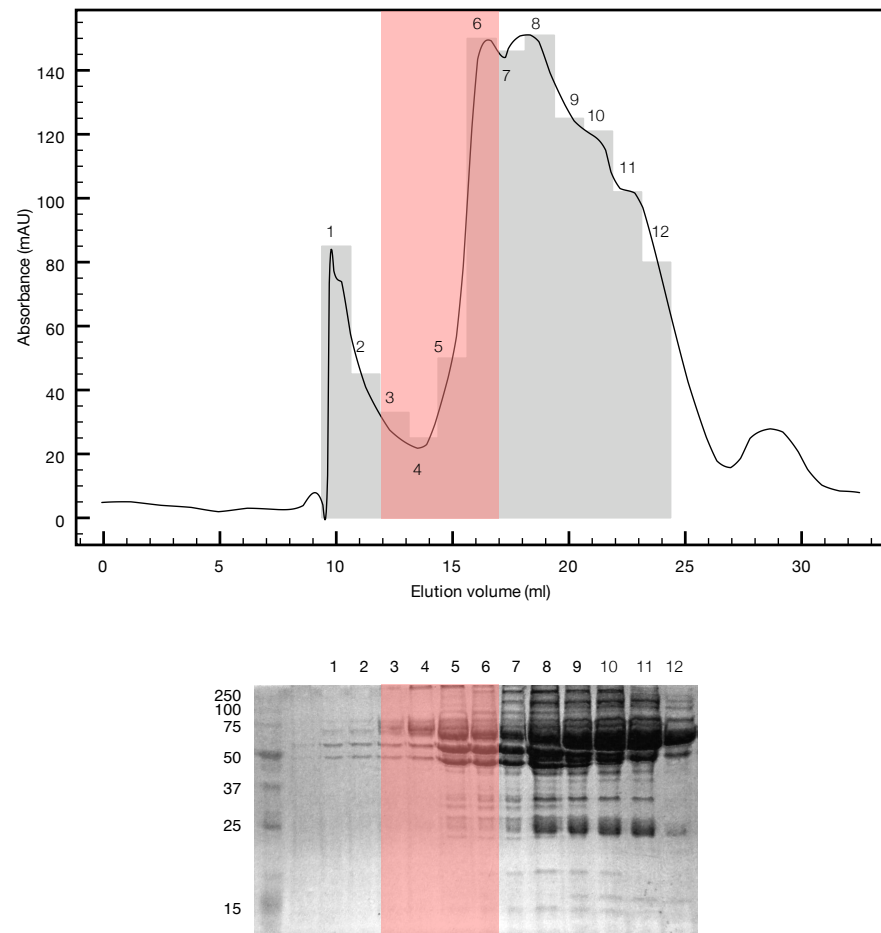


Figure 4.2: C1 purification using Superdex 200. The figure shows a typical chromatogram obtained using the Superdex 200 size exclusion medium and below it the SDS-PAGE gel of the collected fractions. The peak containing C1 is highlighted, and fractions (grey bars) and their corresponding gel lanes are numbered.

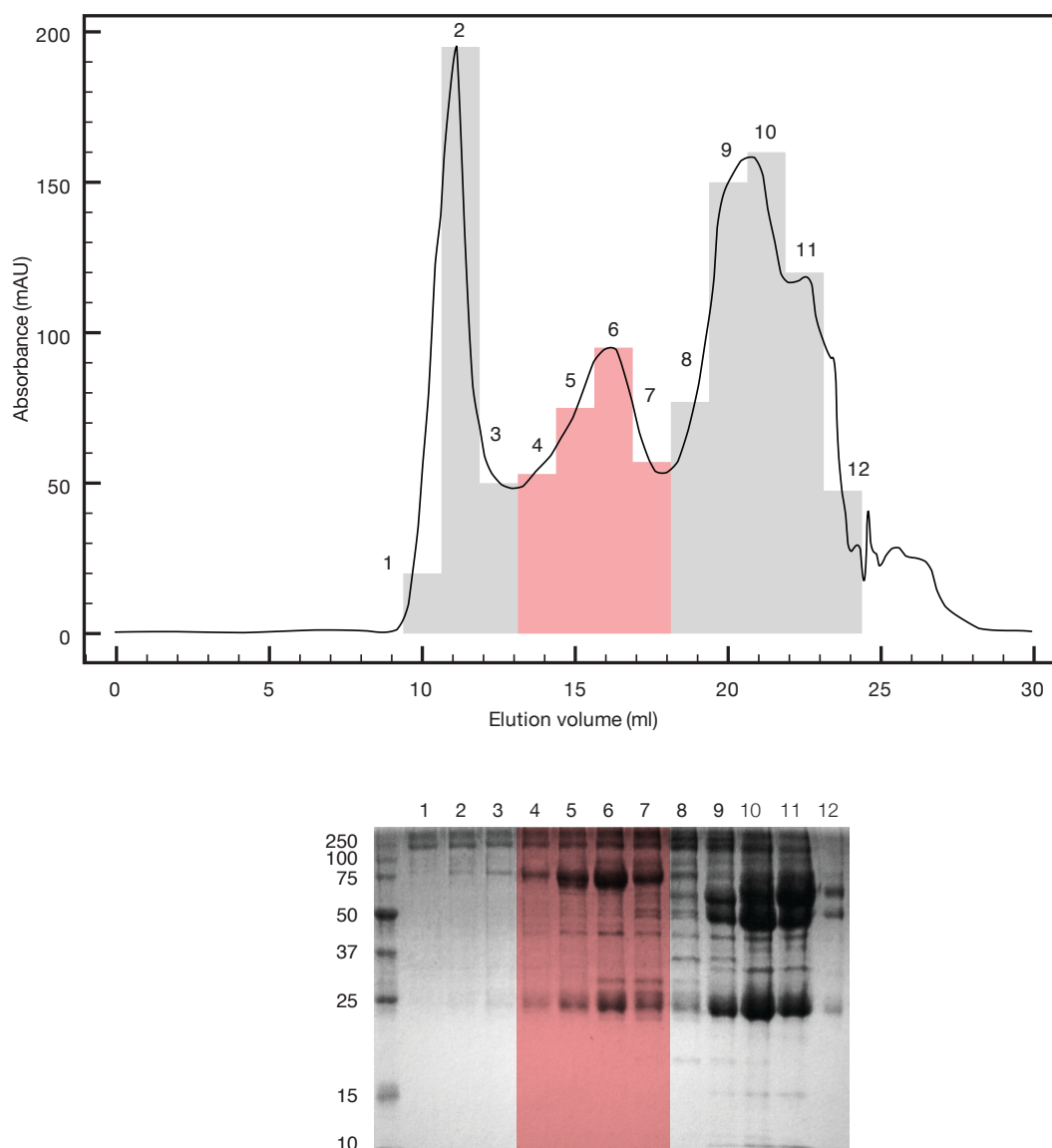


Figure 4.3: C1 purification using Superose 6. This figure shows a typical chromatogram and gel obtained when using Superose 6 as the size exclusion medium. The peak containing C1 is highlighted, and fractions (grey bars) and their corresponding gel lanes are numbered. The results show that a much better separation is achieved when using Superose 6 as the SEC medium. The fractions containing C1 correspond to a single peak in the chromatogram, and show fewer contaminants when analysed by SDS-PAGE.

4.3.2.2 *Effect of buffers and pH on the purification of C1*

The choice of buffer can be a critical factor in protein studies, particularly when crystallisation is desired. Recent results (Long and Yang, 2009) have shown that far from being inert substances, buffers can affect protein dynamics. It is also commonly accepted within the crystallographic community that a particular protein may be more stable, or more amenable to crystallisation, in some buffers than in others. An investigation into a suitable buffer for the purification of the C1 complex was therefore deemed appropriate.

Buffers were selected based on various criteria, notably a high buffer capacity in the physiological range and suitability for downstream applications. Tris was chosen initially, because it is a common buffer for crystallisation trials (see Table 7). However, it is known that Tris binds divalent metal ions in aqueous solution (Hanlon et al., 1966), and it is also unsuitable for cross-linking reactions because the primary amine competes with amide groups on the protein. As an alternative, sodium phosphate was chosen because it is a physiological substance and compatible with downstream applications including cross-linking by EDAC.

Buffer	Entries
Tris	1247
HEPES	331
MOPS	40
Bicine	17
Tricine	6
Phosphate*	1381

Table 7: Buffers in the pH range 6.5–9 and their prevalence in the Biological Macromolecule Crystallisation Database. Data were taken from BMCD 4.0 (Gilliland et al., 1994).

*Phosphate is commonly used as both a buffer and a precipitant in crystallisation.

The choice of buffer (phosphate, Tris, HEPES, MES, MOPS) did not have any noticeable effect on the results of purifying C1 by the method outlined here. This observation was supported by the results of subsequent DLS studies on the purified complex (see Figure 4.4). This is useful information, since the choice of buffer is a key variable in crystallisation trials, and so a purification buffer can be selected that will be compatible with downstream screening methods. For example, while phosphate is very popular, it also results in high numbers of false positive hits when metal salts are included as a crystallisation variable.

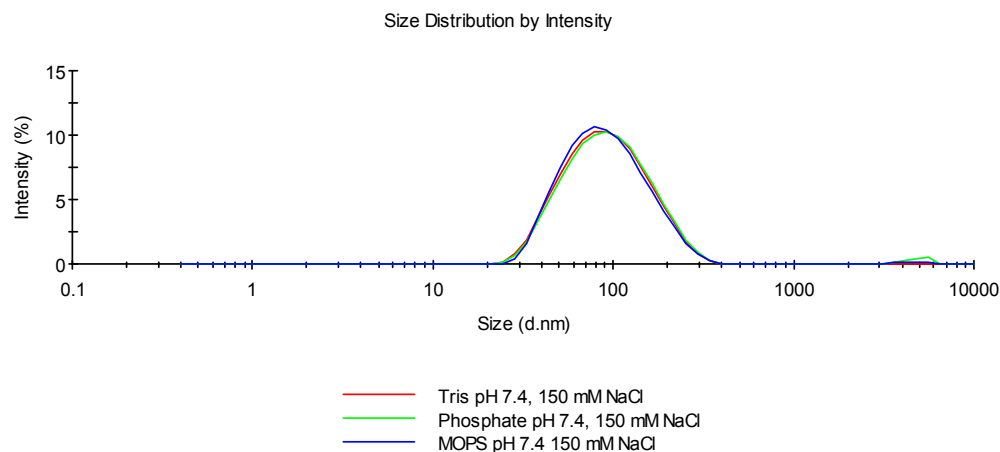


Figure 4.4: The effect of different buffers on C1 purified by SEC. A sample of the purified protein was exchanged into various buffers and analysed by DLS. The samples show a monomodal distribution, with a mean Z-average size of 81.36 nm (± 2.102 nm) and a mean Pdl of 0.212 (± 0.012 nm). This confirms that different buffers do not significantly affect the size of the C1 complex as determined by DLS.

While the buffers themselves may be inconsequential for the purification of C1, pH affects the function of C1 and therefore may alter the structure of the complex. For this reason an investigation was undertaken into the effect of different pH values on the purification of C1. The function of the C1 complex has previously been reported to be greatly affected by pH (Lepow et al., 1958), having optimum activity around pH 7.5,

with a sharp decrease in activity below pH6.5 and above pH8.5. At present the literature does not indicate whether this is due to conformational or chemical changes that render the complex inactive or whether it is due to disruption of the complex.

In light of this, it was decided that the purification of C1 would be performed at physiological pH in order to maximise the amount of material that could be obtained from each PEG precipitation. Since pH affects C1 activity, controlling the pH during the purification procedure could add a valuable level of control to minimise C1 activity if so desired.

A further point to note was that calcium was included as a variable when testing different buffers, and no determinable difference was detected in the chromatogram when the purification was carried out with or without calcium. This suggests that any calcium ions are tightly bound to C1 and that the complex does not require the addition of further calcium during purification.

4.3.2.3 Effect of ionic strength on the purification of C1

In addition to being sensitive to pH, human C1 has also been shown to respond to variations in ionic strength, showing a decrease in activity below 0.1 and above 0.15 (Lepow et al., 1958). In another study, guinea pig C1 was also shown to be sensitive to salt concentration (Colten et al., 1968). Here again, the authors propose that the decrease in activity is due to dissociation of the complex at high ionic strength.

Various salt concentrations (0, 150 and 500mM) were explored for their effect on the elution profile of the crude C1 extract. As with the investigation into the effect of pH, equal amounts of PEG-precipitated proteins were reconstituted in buffers of different ionic strengths and the samples were then separated by size exclusion chromatography. The resulting chromatograms and the corresponding SDS-PAGE gels are shown in Figure 4.5.

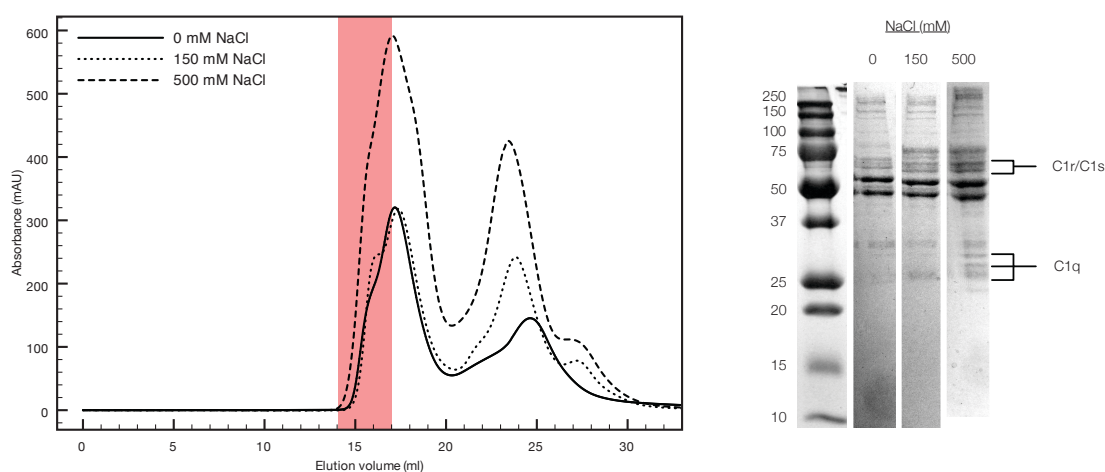


Figure 4.5: Effect of varying salt concentrations on the recovery of C1 from the precipitate. Equal amounts of PEG-precipitated proteins were reconstituted in 10mM MES, pH7.4, 2mM calcium chloride, containing either 0, 150 or 500mM NaCl and separations were performed in the same buffer. The C1 peak in the chromatogram is highlighted and the corresponding SDS-PAGE gel for each salt concentration is shown. The results show that considerably more material is recovered using 500mM NaCl in the re-suspension buffer than when no salt or physiological salt levels are present.

From both the chromatograms and the gel it can clearly be seen that re-suspending in a buffer containing 500mM sodium chloride results in an increased amount of C1 recovered from the precipitated pellet. However, changing the salt concentration in the range tested does not appear to affect the purity or overall integrity of the complex. Despite the apparent advantages of using high salt concentrations during the extraction, further preparations were made using physiological salt concentrations during the purification procedure, in order to obtain purified C1 in its native state. However, if 500mM sodium chloride does effectively inhibit C1 action, this could be a useful adaptation to the purification procedure to prevent any undesired reactions.

4.3.2.4 *Choice of starting material: plasma or serum*

As can be seen from Figure 4.5, the fractions containing C1 that are obtained after an initial gel filtration run contain numerous contaminating proteins. During the course of several preparations it was observed that the purified fractions turned gelatinous upon storage at 4°C. A possible explanation was that residual clotting factors were being co-purified with C1 and that the calcium in the purification buffer was inducing clotting. To test this hypothesis, the purification was repeated using plasma that had been depleted of clotting factors by incubating with 10–50mM calcium overnight at room temperature and removing the resulting clot by centrifugation. The resulting fractions remained liquid upon storage, although some precipitation was observed after several weeks. This suggests that the majority of clotting factors were removed by the treatment, and indeed the resulting fractions were much cleaner when examined on SDS-PAGE gels. Since there was an excess of calcium in the serum, no further calcium was added to the purification buffer.

4.3.2.5 *Re-suspension volume*

The concentration of C1 in normal human serum is 1.8×10^{-7} M (Ziccardi and Cooper, 1977). This equates to 140 mg/l, so in a typical preparation using 50 ml of serum we would expect to recover approximately 7 mg of purified C1. From a single extraction, the amount of protein in the pooled fractions after SEC is around 1.5 mg. Since there are still considerable amounts of other contaminating proteins, we can estimate that less than 1.5 mg of C1 is recovered of a potential 7 mg, or a yield of <20 %.

There are several possibilities to explain this low yield. One explanation is that not all of the C1 present in serum precipitated in the PEG-precipitation step. However, the original purpose of this method was to prepare C1-deficient serum (Neoh et al., 1984) and therefore it is unlikely that a significant quantity of C1 remains in solution. Another explanation is that not all of the precipitated C1 was re-suspended and that some remained in the pellet of precipitated proteins. This was tested by re-extracting the same pellet to see if more C1 could be recovered and the results are shown in Figure 4.6.

It was found that equivalent amounts of high-molecular weight material could be sequentially extracted from PEG-precipitated human serum. The purification protocol was therefore modified so that the PEG-precipitate was re-suspended in 1/10 the original volume of buffer. This modification obviated the need to use high salt concentrations to increase the yield of protein recovered from the precipitate.

In summary, I have demonstrated that C1 can be partially purified quickly and easily from human plasma by a novel two-step purification. A volume of plasma was calcified to remove the majority of clotting factors and the resulting serum was then incubated on ice with the addition of PEG 6000 to 3.5 % in order to precipitate high-molecular-weight proteins including C1. The precipitate was recovered by centrifugation and re-suspended in 1/10 the original volume of buffer (Tris, HEPES, MES, MOPS) at

physiological pH and salt concentration, with or without physiological levels of calcium present in the extraction buffer. This crude extract was purified using a Superose 600 size exclusion matrix, resulting in purified C1.

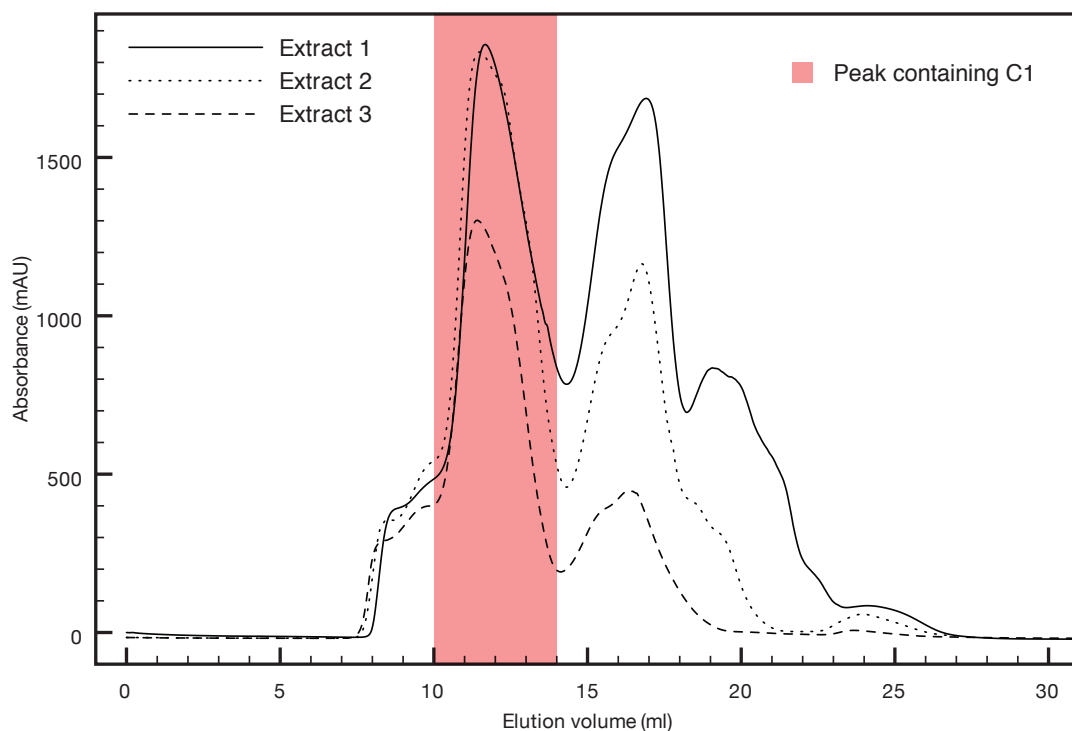


Figure 4.6: Multiple extractions of PEG-precipitated proteins. The pellet obtained from the PEG precipitation step was re-suspended in 1 ml of 10 mM Tris, pH 7.4, 150 mM sodium chloride and any remaining precipitate was removed by centrifugation. A second extraction was then performed on the remaining precipitate, and a third in a similar manner. The figure shows that similar amounts and distributions of high-molecular-weight material were recovered from each extraction. After approximately five extractions, most of the material in the pellet had been recovered.

4.3.3 Summary: Purification by SEC

The optimised method for purifying C1 by SEC is as follows: a volume of frozen human plasma is thawed and clotted with 50mM calcium chloride. The resulting serum is precipitated with 3.5% PEG 6000 for 30 minutes on ice and the precipitate is then re-suspended in 1/10 the original volume of buffer (10mM Tris, pH7.4, 150mM NaCl). Any remaining precipitate is removed by centrifugation and filtration through a 0.2 µm filter. The clarified sample is loaded onto a Superose 6 10/300 GL Tricorn™ high performance column (GE Healthcare, UK) and eluted at 0.5ml/min in the same buffer. Tris may be replaced with another suitable buffer if required.

4.4 Characterisation of C1 purified by SEC

In order to assess whether or not the purification has been successful, the eluted fractions were subjected to analysis by a number of complementary methods. Since the aim of this project is to obtain native C1 in a form suitable for use in crystallisation trials, the techniques employed were chosen so as to determine as clearly as possible whether the purified product obtained is indeed C1 and whether it is pure and structurally homogeneous. This following section describes the various ways in which the results of the purification procedure were evaluated and presents evidence to support the view that the goals set for this project have been adequately fulfilled.

4.4.1 Characterisation of SEC-purified C1 by electrophoresis

One of the most useful methods for assessing the results of the C1 purification was polyacrylamide gel electrophoresis under denaturing conditions, but C1 was also analysed by various other electrophoretic methods. The results of these investigations are presented in the following section, according to the technique employed.

4.4.1.1 *Native agarose gel electrophoresis*

As explained in Chapter 2, native electrophoresis is a method that separates molecules based on their mass/charge ratio. It is gentle enough that interactions between proteins can be maintained and is therefore suitable for the purpose of examining the assembled C1 molecule. Owing to the large size of the complex — around 800kDa — the separation could not be performed in polyacrylamide gels, as is usual for protein separations. Instead, agarose gels were used, as these permit larger pore sizes and are therefore better suited to separating larger molecules. Samples of oxidised and reduced C1, as well as marker proteins thyroglobulin (669kDa, pI 4.6), ferritin (440kDa, pI 4.4) and catalase (232kDa, pI 5.6) were prepared as described on page 126. The proteins were visualised with Coomassie and the results are shown in Figure 4.7.

The results of the native separation show that the purified C1 migrates as a single, high-molecular-weight band of approximately 750kDa. This is consistent with the published molecular mass of the C1 complex at 790kDa. Upon incubation with the reducing agent BME, C1 appears to split into a fragment of approximately 190kDa and a broad band of high-molecular-weight material. The mass of the smaller fragment corresponds roughly to that of a dimer composed of C1r and/or C1s — at around 150 kDa — and the larger matter is presumably the intact C1 complex. Reduction of the disulphide bonds between sub-components therefore apparently leaves the majority

of the molecules intact, with a small proportion forming dimers composed of C1r, C1s or both. It is possible that other forces hold some parts of the complex together under the reducing but non-denaturing conditions during the electrophoresis run.

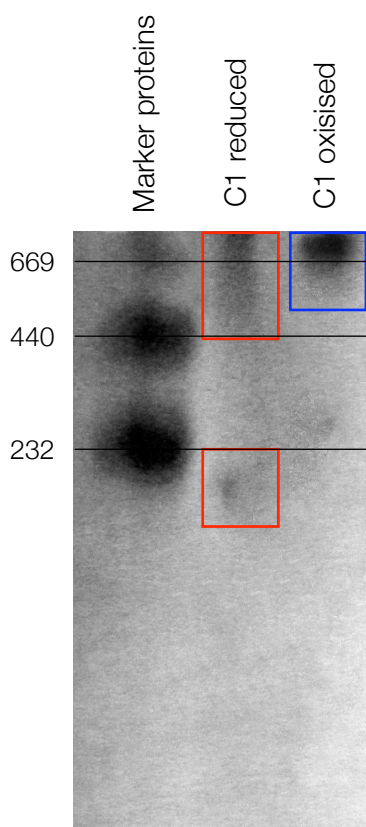


Figure 4.7: Native agarose gel electrophoresis of purified C1. Molecular weight marker proteins: thyroglobulin (669kDa, pI 4.6); ferritin (440kDa, pI 4.4) and catalase (232kDa, pI 5.6). The predicted pI of C1 is 7.3. Purified C1 was reduced by incubation with BME and the gel appears to show two bands (red): one high molecular-weight-band between 440–669kDa, and one low molecular weight band < 232kDa. Purified native C1 (oxidised) migrates as a single band (blue) of over 669kDa, which corresponds to the expected molecular weight of the complex.

Native electrophoresis offers comparatively low resolution compared with other methods: the diffuse bands make the resulting gels difficult to interpret. Nonetheless, since it appears to be non-disruptive to the C1 complex and has recently been

suggested as a highly efficient approach to studying the structure and function of metalloproteins (Kastenholz, 2007), native electrophoresis remains a useful tool in the investigation into the structure of complement component C1.

4.4.1.2 SDS-PAGE

The method of choice for verifying the purification of C1 was SDS-PAGE, since it is quick and simple to perform, and offers a high resolving power. Furthermore, the technique is denaturing and thus allows the detection of the individual C1 components as well as any contaminating substances that elute together with C1 during chromatographic separations. Thus, SDS-PAGE is complementary to other techniques used here — such as SEC, native PAGE and DLS — as it reveals information about the individual polypeptide chains in the sample. An illustration of the results of the purification of C1 by SEC is shown in Figure 4.8. Proteins which are clearly visible in the crude extract have been removed by the purification and the bands corresponding to the various C1 sub-components can clearly be seen. These are barely visible in the impure starting material, indicating that C1 has been enriched by the purification procedure.

When purified C1 is prepared under reducing and denaturing conditions, the complex breaks apart into its individual sub-components. These can be seen as a double band of approximately 80kDa, containing C1r and C1s, and a triplet of bands at around 25kDa, which correspond to the A, B and C chains of C1q. A small amount of very high-molecular-weight (>250 kDa) material is also present, which may be aggregated C1 or other contaminants. For comparison, a published image of reduced and non-reduced, native and activated C1 is shown in Figure 4.9. The results under reducing conditions are in good agreement with the C1 purified here, as shown in Figure 4.8, with the characteristic pattern of ~ 80 kDa and ~ 25 kDa bands.

Under denaturing, but non-reducing conditions, a single prominent band that does not react with silver stain is seen, with a molecular weight greater than 250kDa (Figure 4.8). This is in contrast to the published results (Medicus and Chapuis, 1980), which show bands corresponding to ~ 50 kDa for C1q A–B and C–C chain dimers and ~ 80 kDa for C1r and C1s.

One notable difference between the published protocol and that presented here is the inclusion of a TCA precipitation step prior to electrophoresis; the other difference being the method of purifying C1. Since the acid precipitation step employed by Medicus and Chapuis simply serves to concentrate the proteins present in the sample and the remaining acid is neutralised in the protocol described, it seems unlikely that this is the cause of the observed differences. So perhaps they can be explained by the different methods of preparation.

In the published method, C1 is purified from serum by affinity chromatography, using immobilised Ig-G to bind and capture the complex, which is subsequently eluted with high salt buffer (0.4M NaCl). These steps may in some way facilitate the disruption of the complex by SDS, while the C1 isolated by the procedure described here remains resistant to denaturation by the detergent. Alternatively, C1 purified by this method is for some reason prone to aggregation, and the observed high-molecular-weight band may simply be large denatured protein aggregates. It is interesting that a high-molecular-weight band also appears under non-reducing conditions in the paper by Medicus and Chapuis, 1980 (Figure 4.9).

With respect to the ineffectiveness of silver staining on non-reduced C1, a possible explanation lies in the mechanism by which silver atoms accumulate at the protein surface. In order for the stain to develop, a few soluble silver ions must be reduced to silver atoms by reducing sites on proteins. These microscopic deposits then act as a nucleus from which silver grains grow, visible to the naked eye as dark bands. The colour of the bands varies with the size of the grains (Merril et al., 1988). It is possible that C1 purified by PEG precipitation and SEC, whether in its native state or present as aggregates, presents few or no reducing sites and therefore does not produce a visible stain using silver. Without additional information, it is not possible to draw any further conclusions about why C1 as purified here — in contrast to previous reports — migrates with a molecular weight of greater than 250kDa by non-reducing SDS-PAGE and does not respond to silver staining after staining with Coomassie.

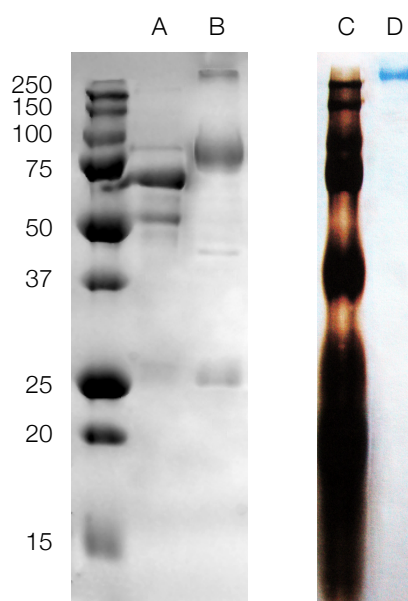


Figure 4.8: Comparison of PEG precipitated proteins and purified C1 under reducing and non-reducing conditions by SDS-PAGE. A. Crude C1 from PEG precipitation of human serum. B. C1 purified by SEC under reducing conditions. C. Marker proteins stained with Coomassie, followed by silver staining. D. C1 purified by SEC under non-reducing conditions, and stained with Coomassie. Subsequent staining with silver stain did not produce any further reaction.

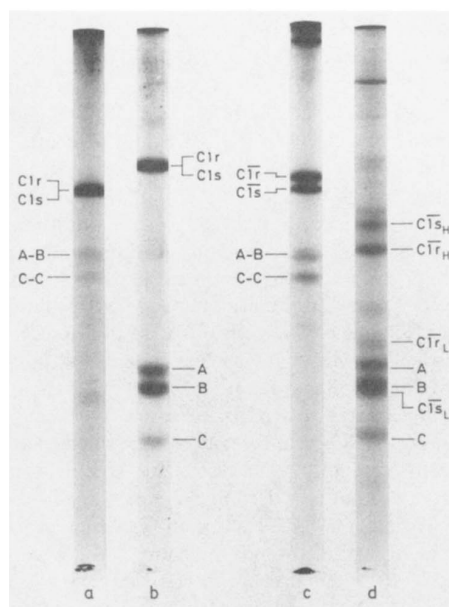


Figure 4.9: SDS-PAGE analysis of native and activated C1 on 7% SDS-polyacrylamide gels. A. Native C1 under non-reducing conditions. B. Native C1 reduced. C. Activated C1 under non-reducing conditions. D. Active C1 reduced. Taken from Medicus and Chapuis, 1980.

4.4.2 Peptide mass fingerprinting of SEC-purified C1

The purpose of this investigation was twofold: firstly, peptide mass fingerprinting will allow the various components of the C1 complex to be identified in the corresponding gel sections, thereby allowing us to confirm that the purification has been successful; secondly, the analysis will give an indication of the major contaminating proteins present within each sample. This information is useful to monitor the efficacy of the purification and can also shed light on any proteins that may be co-purifying specifically with the C1 complex.

Another possible method to determine the presence of a specific protein, for example C1r, C1s or C1q, would be a Western blot. In this technique, proteins are electroblotted from the SDS-PAGE gel onto a porous membrane, which is then incubated with an antibody specific to the protein in question. However, anti-C1 antibodies are expensive, especially considering the need for a separate antibody specific to each subcomponent and the limited availability of suitable commercial antibodies. Furthermore, Western blotting can give only very limited information about the purity of the sample. Using a mass spectrometric approach circumvents these obstacles, and provides detailed quantitative information about the sample composition.

After electrophoresis on SDS-PAGE gels, the C1 complex purified was sent for analysis at the Sussex Proteomics Centre (UK). The various bands observed were excised and submitted for peptide mass fingerprinting in order to determine the identities of the proteins contained within each band, as shown in Figure 4.10. Two separate analyses were performed, using slightly modified gel preparation methods as described on page 127. The results of the two experiments are compared in Table 8.

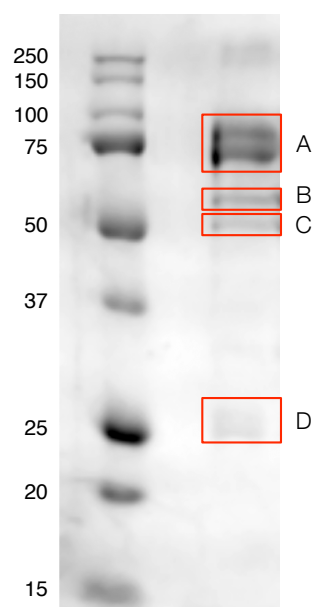


Figure 4.10: Gel submitted for MS analysis of SEC-purified C1. The gel sections outlined in red (A – D) were excised and submitted for in-gel trypsinolysis followed by peptide mass fingerprinting at the Sussex Proteomics Centre.

	Method 1		Method 2	
Human proteins	126	(100 %)	70	(100 %)
Keratins	42	(33 %)	12	(17 %)
Immunoglobulins	32	(25 %)	30	(43 %)

Table 8: Comparison of two different gel preparation methods for peptide mass fingerprinting. In the first method, samples were prepared under normal laboratory conditions, while the second method involved the use of a laminar flow hood to reduce exposure to contamination by keratins from dust, skin and hair.

A major problem with this approach was contamination of the samples, both from extraneous sources and due to the inherent complexity of the samples. When preparing biological samples for mass spectrometry, contamination with keratin from hair, skin and dust is almost unavoidable. In the initial experiment all glassware was thoroughly cleaned, all reagents were freshly prepared, gloves were worn throughout the

electrophoresis run and the remaining sample preparation, and care was taken to minimise exposure of the gel to dust during staining and handling. However, of all 126 human proteins returned in the search, almost 60% were contaminants or high-abundance serum proteins, with one in three proteins being contaminating keratins.


The method of preparing the samples was subsequently modified so that the gel pieces were excised under a laminar flow hood in order to further reduce the exposure of the samples to airborne contaminants. This method performed better, with only 17% of the total proteins identified as keratin. However, the proportion of immunoglobulins jumped from 25% to 43% in the second experiment. These are the most abundant proteins found in human serum, after albumin (Diem and Lentner, 1975), and so it is not surprising that they are prominently represented in the search results. IgG has a serum concentration of 8–18mg/ml, compared to 0.18mg/ml for C1, making it more abundant than C1 by a factor of around 100.

A large proportion of the proteins identified in the peptide search were contaminating keratins or immunoglobulins, but low-abundance serum proteins could also be identified. The combined results of the two experiments are shown in Table 9. The returned search results were automatically filtered to exclude non-human proteins, keratins and immunoglobulins, and the list of proteins was manually refined to eliminate human non-serum proteins. Finally, abundant plasma proteins such as serum albumin, lipoproteins, fibrinogen, alpha-2-macroglobulin and alpha 1-antitrypsin were also manually excluded. These proteins are likely to be present in serum-derived samples owing to their sheer prevalence, as they make up about 8% of total serum protein (the majority of the fibrinogen being removed during the preparation of serum from plasma).

All elements of the C1 complex were identified in the SEC-purified sample, confirming that the intact C1 has been successfully purified from human serum by the method outlined in this work. In addition, several other complement-related proteins were identified in the purified sample, including C3, C4 and C4BP, as well as SAP and various other proteins thought to be involved with the complement system.

Table 9: Results of peptide mass fingerprinting on C1 purified by size exclusion chromatography. The table shows the various C1 subcomponents and other complement-related proteins identified in the sample. Molecular weights determined from protein sequence databases are given as a guide only. Further information is given in the comments. Table continued overleaf (sequence coverage data not available, protein probability scores are shown instead).

Sample	Identified proteins	Accession no.	MW	Sequence coverage	Comments
A (75 kDa)	C4b-binding protein alpha chain precursor	C4BP_HUMAN	67 kDa		6-7 α chains: 70 kDa ¹
	Complement C4-A precursor	CO4A_HUMAN	193 kDa		β chain: 75 kDa ¹
					α chain: 97 kDa ¹
D (25 kDa)	Vitamin K-dependent protein S precursor	PROS_HUMAN	75 kDa		Circulates in complex with C4BP ²
	Complement C3 precursor	CO3_HUMAN	187 kDa		β chain: 75 kDa ¹
	Complement C4-A precursor	CO4A_HUMAN	193 kDa		α chain cleaved into C4a (8 kDa) and C4b (85 kDa); further processed into C4d (45 kDa) ¹
	C4b-binding protein alpha chain precursor	C4BP_HUMAN	67 kDa		C-terminal fragment?
	C4b-binding protein beta chain precursor	C4BB_HUMAN	28 kDa		1 or 0 β chains: 25 kDa ¹
	Complement factor I precursor	CFAI_HUMAN	66 kDa		Light chain: 28 kDa ³
	CD5 antigen-like precursor	CD5L_HUMAN	38 kDa		C-terminal fragment?
	Complement C1q subunit C precursor	C1QC_HUMAN	26 kDa		N-terminal fragment?
	EF-hand calcium-binding domain-containing protein 3	EFCB3_HUMAN	50 kDa		
	Serum amyloid P-component precursor	SAMP_HUMAN	25 kDa		

Sample	Proteins identified	Accession no.	MW	Probability (protein)	Score	Comments
B						
(60 kDa)	C4b-binding protein alpha chain precursor	C4BP_HUMAN	67 kDa	6.83E+01	88.28	N/A
	Complement C1r subcomponent precursor	C1R_HUMAN	80 kDa	4.98E+01	70.21	Heavy chain: 51 kDa ³
						
C						
(50 kDa)	Complement C4-A precursor	CO4A_HUMAN	193 kDa	5.97E+01	30.24	N/A
	C4b-binding protein alpha chain precursor	C4BP_HUMAN	67 kDa	4.48E+01	40.20	N/A
	C4b-binding protein beta chain precursor	C4BB_HUMAN	28 kDa	2.51E+01	30.22	N/A

1. Morgan, B. P. (ed) Complement methods and protocols. (Humana Press, Totowa, NJ, 2000).
2. Webb, J. H., Blom, A. M. & Dahlback, B. Vitamin K-dependent protein S localizing complement regulator C4b-binding protein to the surface of apoptotic cells. J Immunol 169, 2580-2586 (2002).
3. The Universal Protein Resource (UniProt) in 2010. Nucleic Acids Res 38, D142-8 (2010).

4.4.3 Dynamic light scattering of SEC-purified C1

The acquisition of a dynamic light scattering instrument part-way into the project offered another complementary means for analysing the purified fractions. As described in section 2.3.3, dynamic light scattering gives information about the size distribution of particles in solution. Thus it differs from PAGE and SEC as an analytical tool in that it gives us information about more than simply the size of the molecules under scrutiny.

The technique offers other advantages too. The fact that it is non-destructive, and can accommodate small volumes makes it ideal for use when there is only limited sample available, or when the sample is required for further downstream applications. Additionally, any changes in the size distribution of the sample can be monitored in real time. Thus, time course experiments to test the effect of temperature or the addition of various substances to the solution can be performed and monitored directly. Furthermore, DLS can be used to estimate the shape of macromolecules and make predictions about their ability to crystallise under given conditions.

The combination of the above factors makes DLS a powerful technique for studying macromolecules from a structural biology point of view. As with SDS-PAGE, dynamic light scattering was used extensively throughout this work, and individual results are presented in the relevant sections. However, an analysis of C1 purified by SEC is shown in Figure 4.11, p. 171.

The diameter of the assembled C1 molecule has been proposed to be around 40nm (Perkins, 1985). For C1 purified by size exclusion chromatography, we see a broad, convoluted peak in the DLS intensity plot in Figure 4.11. Multimodal analysis gives two conjoined peaks, and is more likely to reflect the nature of this polydisperse sample accurately. The first peak is centred at approximately 45nm, which is very close to the proposed diameter of assembled C1 (40nm). There is a slightly smaller second peak at around 180nm. Of course the 45nm particles are more numerous than the larger ones present in the sample (for an overview of interpreting DLS data, see section

2.3.3.2, p. 63). The width of the distribution indicates that there is likely to be a range of particle sizes, possibly reflecting various different contaminants or states of protein association.

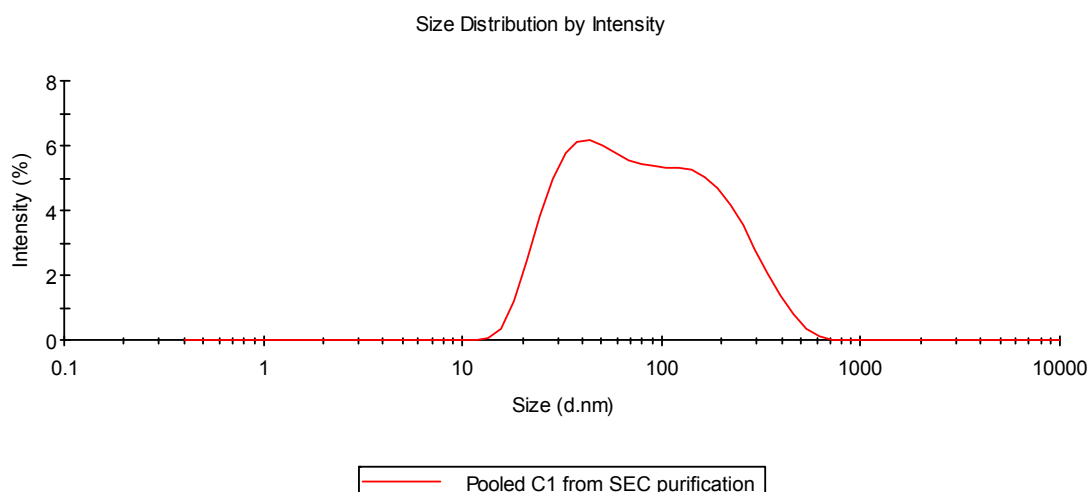


Figure 4.11: Size distribution of pooled fractions SEC-purified C1 by intensity. The protein was purified using the optimised purification procedure, and fractions corresponding to the C1 peak were pooled and analysed by DLS. The plot shows two conjoined peaks centred at approximately 45nm and 180nm. The smaller particles outnumber the larger particles, since larger particles scatter more strongly in proportion to their size, and there is likely to be a range of particle sizes and association states.

When each fraction comprising the C1 peak from the optimised SEC purification procedure was assayed individually, we can see that the size distribution varies between fractions (see Figure 4.12, p. 172). The first fraction contains a broad peak at around 100nm, which is skewed towards the lower end of the range. In the second fraction, the size distribution becomes bimodal. The later fractions show two distinct peaks: a large peak at 30nm and a smaller peak at 180nm. This indicates that the C1 peak purified by the optimised SEC procedure contains a population of particles of around 30nm, with a smaller amount of larger particles that steadily decreases with increasing

fraction number. Again, the size of the smaller particles is close to the proposed size of the intact C1 complex (40nm), taking into account shape effects and the limitations of the technique.

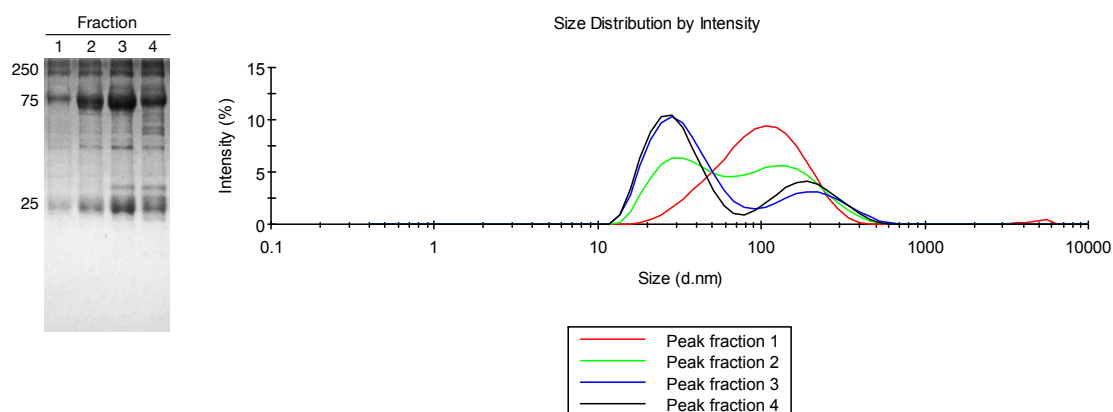


Figure 4.12: Size distribution by intensity of individual fractions of the C1 peak purified by SEC. The protein was purified using the optimised purification procedure, and fractions corresponding to the C1 peak were analysed individually by DLS: the corresponding SDS-PAGE gel is also shown (left). The intensity plot shows how the distribution of particle sizes varies between fractions, around two peaks at around 25nm and 180nm. In the first fraction, the larger particles appear to dominate, with a large peak at around 100nm. As the fraction number increases, the peak at 25 nm decreases.

4.4.4 Transmission electron microscopy of SEC-purified C1

TEM was used to examine the purified complex with a view to ultimately obtaining improved low resolution structures by cryo-electron tomography. C1 is an ideal candidate for this approach thanks to its large size and the symmetry present in the assembled complex. The TEM sample preparation was carried out according to a modified version of the method of Strang et al., 1982, which uses the zero length cross-linker EDAC to fix the interacting C1 subunits. This water soluble carbodiimide causes carboxyl groups to react with primary amines, resulting in an amide bond. Closely interacting proteins may thus be covalently linked, since any exposed carboxyl groups will bind to nearby amine side chains. Figure 4.13 shows typical views of the cross-linked C1 sample.

Strongly negatively stained features of approximately 50nm and 200nm were observed in all samples tested, as well as numerous diffuse features of approximately 50nm. The strongly stained features are comparatively few, while the diffuse features are the most numerous of the features observed by TEM. The size of these ultra-structural features also corresponds to the distribution of particle sizes as determined by DLS, suggesting that the observed features are real and not artefacts. C1 molecules have been shown to appear diffuse when visualised by this method (see Figure 4.14 and Strang et al., 1982). It thus seems likely that the abundant diffuse features observed here are C1 molecules. However more work is required to fully resolve the structural detail of the sample.

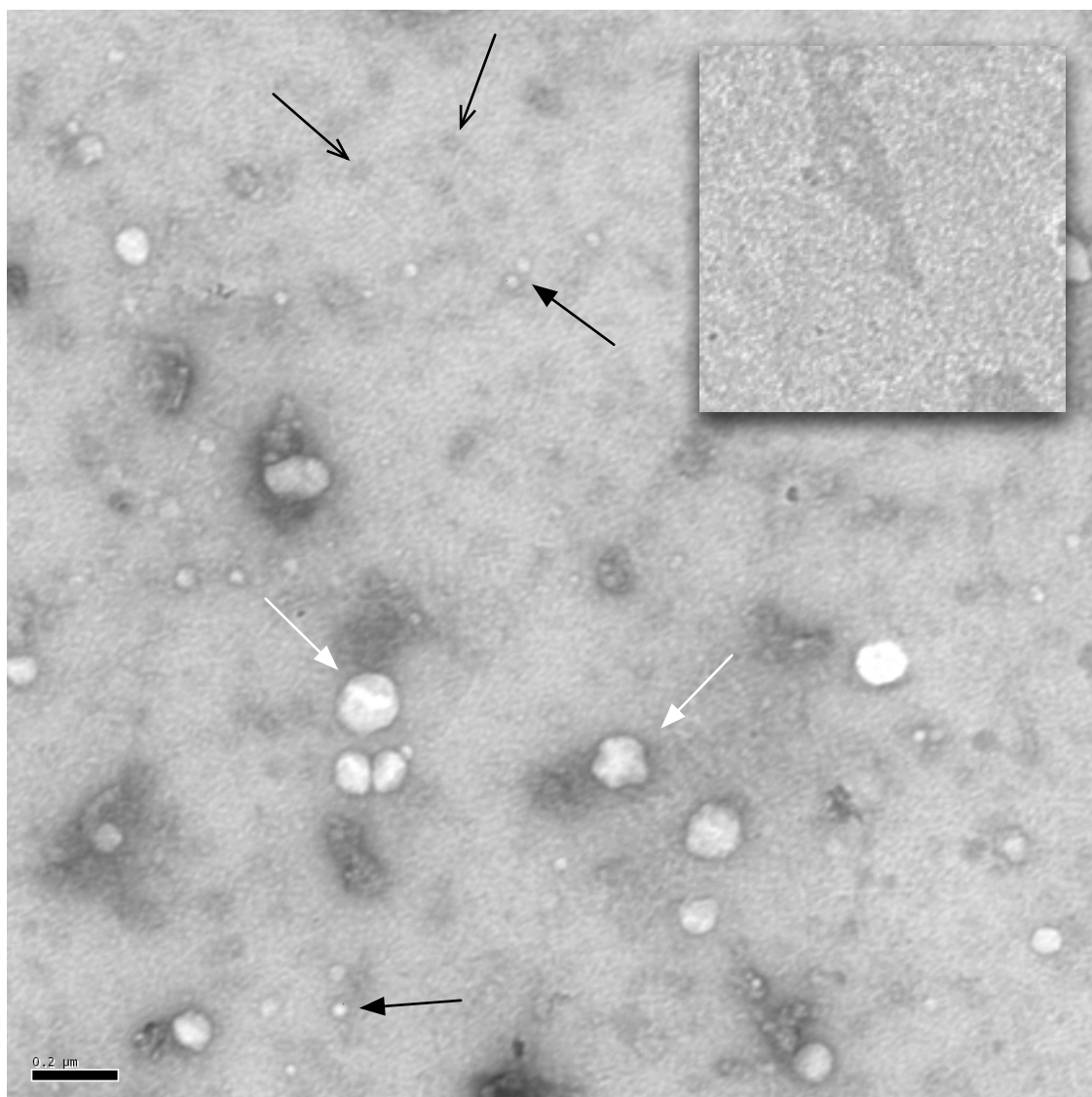


Figure 4.13: Cross-linked C1 analysed by TEM. A sample of purified C1 was cross-linked using the zero length cross-linker EDAC and visualised by negative staining at 10000 × magnification. Inset: negative control sample with no protein present. Globular negatively stained features of approximately 50nm (black arrows) and 200nm (white arrows) were seen in the sample, which were not observed in negative controls. In addition, numerous diffuse features were observed in the C1 sample, again approximately 50nm in diameter (open arrows). The particles observed by TEM correspond to the size distribution of particles as determined by DLS.



Figure 4.14: Cross-linked C1 (for comparison). Cross-linked C1 at $290\,000\times$ magnification, taken from Strang et al., 1982. The C1 molecules appear diffuse, with little contrast against the background and surrounded by dark halos.

4.4.5 Storage of SEC-purified C1

Since complement component C1 spontaneously activates at 37°C (Ziccardi, 1983), it is possible that the purified protein may degrade on storage. For concentrated samples, a floccular precipitate was observed after the samples were stored at 4°C for several days. The precipitate could not be removed by centrifugation but only by filtering the sample through a 0.2 µm filter. To test the effect of different storage conditions, they were compared using DLS to monitor any changes in the size distribution of the sample, which would indicate any degradation or aggregation of the purified protein. Samples of C1 prepared by SEC were analysed by DLS before being stored for 10 days at 4°C or -20°C. The samples were then compared with the original, freshly prepared sample, and the results are shown below in Table 10.

As can be seen, storing the sample at -20°C produced the greatest change in the average size of the particles within the sample. This indicates that purified C1 aggregates or precipitates upon freezing and thawing, which is not uncommon for many proteins. Storage at 4°C also resulted in a less pronounced increase in the diameter of the particles, with the polydispersity of the sample remaining virtually unchanged. Overall, the results show that the composition of the SEC-purified C1 fractions changes on storage and this must be taken into account when performing further experiments and must be borne in mind when interpreting the results.

Conditions	Z-Average (d.nm)	Pdl
Freshly prepared	74.73	0.244
4°C	87.14	0.225
-20°C	96.42	0.229

Table 10: DLS measurements of SEC-purified C1 stored for 30 days under various conditions. The results show that a small amount of aggregation occurs when the purified protein is stored for 10 days at 4°C and -20°C. The precipitate could be removed by filtration.

4.4.6 Summary: Characterisation of C1 purified by SEC

The optimised SEC purification procedure outlined in section 4.3 yields a sample that is highly enriched in C1, with a small amount of contaminants. These results were confirmed by native agarose gel electrophoresis, SDS-PAGE and DLS. Peptide mass fingerprinting suggests that the majority of the contaminants are immunoglobulins and other abundant serum proteins, as well as other members of complement and complement-related proteins. In summary, the method outlined here represents a novel way of obtaining the native C1 complex in purified form.

4.5 SEC-purified C1 under physiologically relevant conditions

The results of the investigations into the effects of ionic strength and pH on the purification of complement component C1 by SEC are described in section 4.3.2, p. 146. However, purified C1 was also studied under various other physiologically relevant conditions using gel filtration chromatography and dynamic light scattering to determine the effect of ionic strength, pH and divalent metal cations on the sample.

Human blood plasma contains a variety of organic and inorganic molecules. These are usually confined to a narrow concentration range, and any variation outside of these values may be an indicator, if not a cause, of disease. Fluctuations in levels of certain substances are also used to convey signals within and between biological systems. For example, nerve impulses are generated by the influx of sodium ions to the firing cell through voltage-gated ion channel proteins embedded in the plasma membrane.

In the search for knowledge about the structure and function of C1, the effect of biologically relevant metal ions cannot be ignored. In addition to gaining insight into the physiology of the complex, an understanding of how salt, pH and various metal ions affect C1 in solution can be of help in the search for conditions that allow crystallisation of the assembled complex. The outcomes of these experiments are detailed below.

4.5.1 The effect of ionic strength on SEC-purified C1

As outlined previously, the function of complement component C1 is affected by varying the ionic strength of the surrounding medium (Lepow et al., 1958; Colten et al., 1968). The authors report a decrease in activity at ionic strengths below 0.1 and above 0.15, which is believed to be caused by dissociation of the complex. While the investigations detailed on page 152 showed that the quality of the purification is not appreciably affected by variations in ionic strength, the purpose of the experiments presented here was to investigate the effect of ionic strength on the solution structure of the C1 complex purified by the novel method developed in this work.

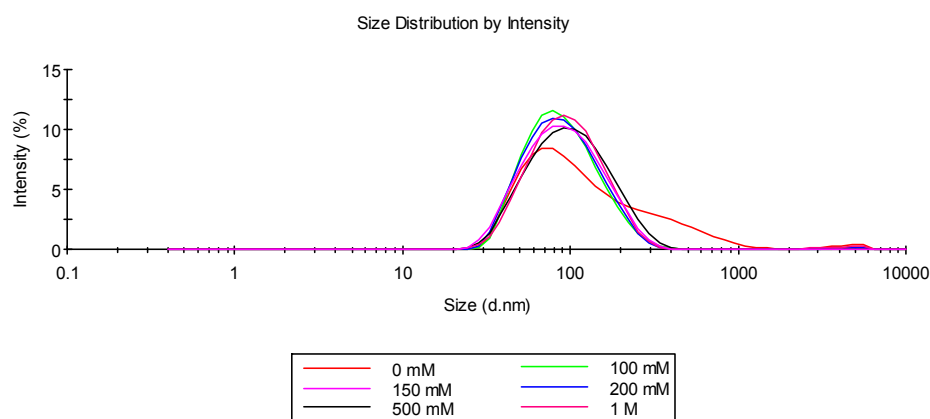


Figure 4.15: The effect of increasing concentrations of NaCl on SEC-purified C1. In the absence of NaCl, purified C1 appears to contain larger aggregates in addition to numerous smaller particles. In the presence of 150–1M NaCl, the size distribution is monomodal. However, the width of the distribution indicates that the sample may contain different populations of particles. These may be contaminants or different states of association of the C1 complex.

When SEC-purified C1 is buffer exchanged into a low ionic strength buffer (10mM Tris, pH7.4), the size distribution of the particles in solution is broad and bimodal, indicating that C1 is associating to form higher-order aggregates (see Figure 4.15). In contrast, when the buffer contains NaCl at 100mM or greater concentrations, the size

distribution becomes narrower and monomodal. A more detailed analysis is shown in Table 11. We can see that three distinct size distribution profiles emerge depending on the ionic strength of the sample buffer. When no additional salt is present in the buffer, the protein population is polydisperse ($\text{PdI} = 0.342$), showing a bimodal distribution with a Z-average size by intensity of around 96 nm. In the presence of physiological salt concentrations (100–200 mM), the particle distribution becomes monomodal (mean $\text{PdI} = 0.199$) and the mean Z-average size decreases to around 80 nm. Finally, at high ionic strength (0.5–1 M salt), the mean Z-average size increases slightly to around 86 nm, while the polydispersity remains essentially unchanged (mean $\text{PdI} = 0.200$).

Range	NaCl (M)	PdI	Z-Average (d.nm)
Low	0	0.342	95.73
	0.10	0.190	80.67
Physiological	0.15	0.205	80.16
	0.20	0.201	80.38
High	0.5	0.205	86.26
	1	0.195	85.29

Table 11: The effect of ionic strength on the size and distribution of SEC-purified C1 by DLS. Three different states can be determined. In a low ionic strength buffer at pH 7.4, the sample is polydisperse with a Z-average size of around 95 nm. In the presence of physiological salt concentrations, the Z-average size decreases to 80 nm and the polydispersity also decreases. At high ionic strength, the Z-average increases again to around 85 nm, while the polydispersity remains the same.

Overall, the results here show that increasing the ionic strength does not cause the C1 complex to disassemble. Instead, at salt concentrations that fall outside the physiological range, complement component C1 appears to have an increased hydrodynamic radius. At low ionic strength, the combination of an increased average size and high sample polydispersity suggest a tendency towards aggregation. However,

at high ionic strength the polydispersity does not parallel the increase in average size. This suggests that the population of molecules is relatively homogeneous and therefore any changes in the apparent size of particles may be attributed to shape effects resulting from conformational changes within the molecule in response to the elevated ionic strength of the surrounding medium.

The results of these investigations on purified C1 also explain why similar amounts of C1 were recovered during purification using buffers of different ionic strengths (section 4.3.2.3, p. 152): if C1 were dissociating in high or low salt buffers we would expect less material to be recovered, but this is not the case. Taken together these observations are consistent with the results published by Lepow et al., 1958 and Colten et al., 1968 and suggest that the reported decrease in activity at ionic strengths below 0.1 and above 0.15 does not result from dissociation of the complex as previously assumed. Instead, the decrease in C1 function must be caused by some other mechanism, for example conformational changes making the catalytic core less accessible or distorting the serine protease domains to render them inactive.

4.5.2 The effect of EDTA on SEC-purified C1

The addition of 8–50mM EDTA is known to dissociate C1 into its individual sub-components (Lepow et al., 1963; Peitsch et al., 1988). In order to determine the effect of EDTA on C1 purified by the optimised SEC procedure (as outlined in section 4.3.2, p. 146), an aliquot of the purified complex was subjected to analytical SEC in the presence of excess EDTA (50mM) in the sample and elution buffer. It was observed that very little protein eluted from the column when EDTA was included in the buffers, indicating that the sample had precipitated to a large extent. As a control, a much smaller quantity of starting material was loaded when using physiological calcium concentrations in the sample and elution buffers. The results are shown in Figure 4.16.

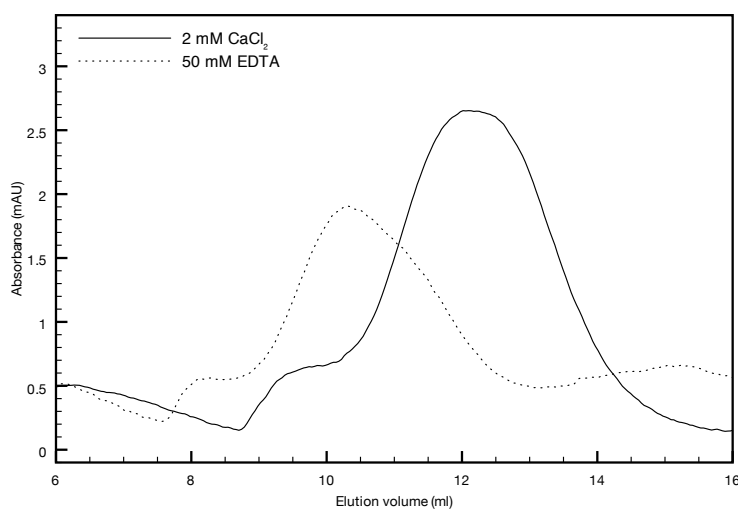


Figure 4.16: The effect of EDTA on SEC-purified C1 by analytical SEC. A sample of purified C1 was subjected to further size exclusion chromatography under physiological conditions (10mM MOPS, pH7.4. 150mM sodium chloride). The buffer additionally contained either 2mM calcium chloride (solid line) or 50mM EDTA (dashed line). Different amounts of sample were loaded on to the column to give a similar amount of eluting protein. When calcium is present, purified C1 elutes as a single peak at around 12ml. In the presence of 50mM EDTA the peak is shifted to around 10 ml, which suggests an increase in molecular weight.

The peak is clearly shifted to the left when EDTA was included in the sample buffer, indicating that aggregates are forming. This would explain the observed decrease in protein eluting from the column, since larger aggregates will be trapped by the in-line filters during the chromatography run. In order to confirm that the protein particles are indeed aggregating and to determine the effect of different EDTA concentrations on SEC-purified C1, the sample was further analysed by DLS. The results of these analyses are shown in Figure 4.17, which clearly shows that increasing concentrations of EDTA cause aggregation of the sample.

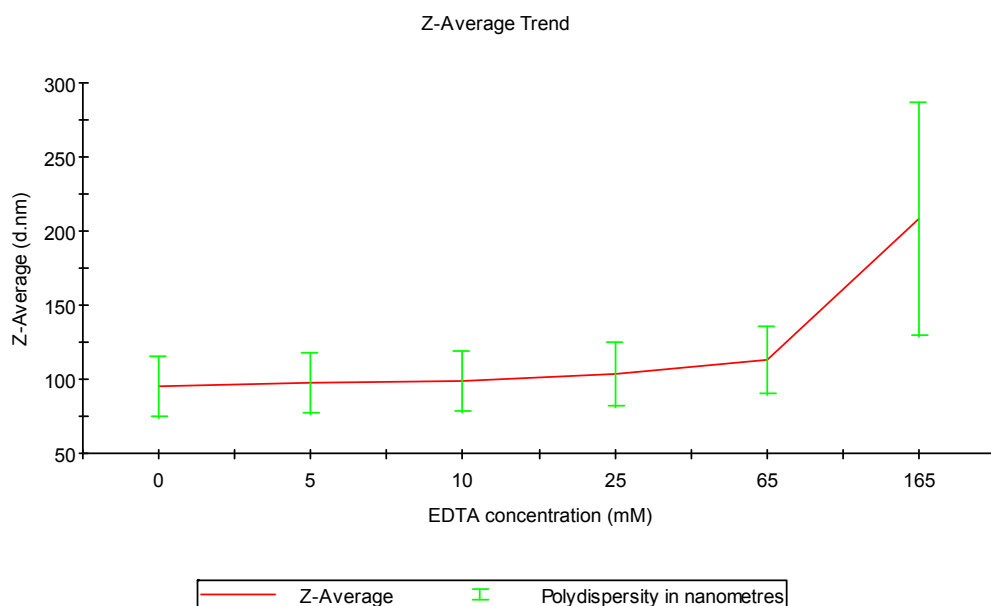


Figure 4.17: The effect of EDTA on SEC-purified C1 by DLS. The C1 complex was purified in MOPS, pH7.4, 150mM NaCl, 2mM calcium chloride and titrated with 100mM EDTA, pH7.4. The plot shows that the Z-average size increases with increasing EDTA concentration, indicating that EDTA induces the aggregation of C1 purified by this method.

4.5.3 Divalent metal cations, pH and SEC-purified C1

Complement component C1 requires calcium for structural stability and functional activity, but other divalent metal cations have also been shown to promote the assembly of a functional complex (Ziccardi, 1983). These studies were carried out at physiological pH and using a range of metals from the first transition period of the periodic table.

Substance	Concentration (approx.)
Total protein	60–80 mg/ml
<i>albumin</i>	35–55 mg/ml
<i>globulins</i>	25–35 mg/ml
<i>fibrinogen</i>	2–4 mg/ml
Sodium	135–145 mM
Potassium	3.5–5 mM
Calcium	
<i>total</i>	2–3 mM
<i>ionised</i>	1 mM
Magnesium	0.5–1 mM
Iron	25–200 μ M
Copper	10–25 μ M
Zinc	10–20 μ M

Table 12: Reference ranges for common substances in human plasma. The typical ranges are given for abundant serum proteins and common metal ions.

Table 12 shows the most prevalent metal ions in human blood plasma, including calcium, magnesium, copper and zinc. Calcium and zinc were shown to promote C1

formation, while magnesium was ineffective and copper was not included in the study. Copper is an essential nutrient and plays an important role in oxidative defence. Excessive or deficient copper levels have been implicated in various diseases and the concentration of copper is apparently increased at sites of infection or inflammation (Uriu-Adams and Keen, 2005). Furthermore, it is generally accepted that the pH drops at local sites of infection or inflammation, which may in turn affect proteins at the site and their interactions. In order to investigate the effect of physiologically relevant cations on C1, the purified sample was analysed by DLS in the presence of calcium, magnesium, zinc and copper at both neutral and acidic pH.

4.5.3.1 SEC-purified C1 at physiological and acidic pH

C1 function is reported to be greatly affected by pH (Lepow et al., 1958): the complex appears to possess optimum activity at physiological pH, with a sharp decrease in activity below pH6.5 and above pH8.5. Whether this is due to conformational or chemical changes that render the complex inactive, or due to disruption of the complex is unknown at present. When the SEC-purified sample was exchanged into a low pH buffer (10mM MES, pH 5.5), no dramatic changes were observed in the size distribution profile, as can be seen from the data in Table 13. This suggests that C1 does not dissociate at low pH and that a decrease in function may be explained by conformational changes within the molecule.

Sample	Z-Average (d.nm)	Pdl
pH7.4	80.16	0.205
pH5.5	83.64	0.203

Table 13: SEC-purified C1 at physiological and acidic pH. The slight increase in Z-average size at pH5.5 compared with pH7.4 may correspond to conformational changes in C1.

4.5.3.2 Divalent metal cations at neutral and acidic pH

a) Calcium, magnesium and zinc at pH 7.4

The iron in human blood is primarily found associated with proteins such as haemoglobin or the iron transport protein transferrin, while copper is mainly bound by ceruloplasmin. Consequently, the most common free divalent metal cations in human blood plasma are calcium, magnesium and zinc. These were investigated for their effects on purified C1 under physiological conditions and the results are shown in Figure 4.18 and Table 14.

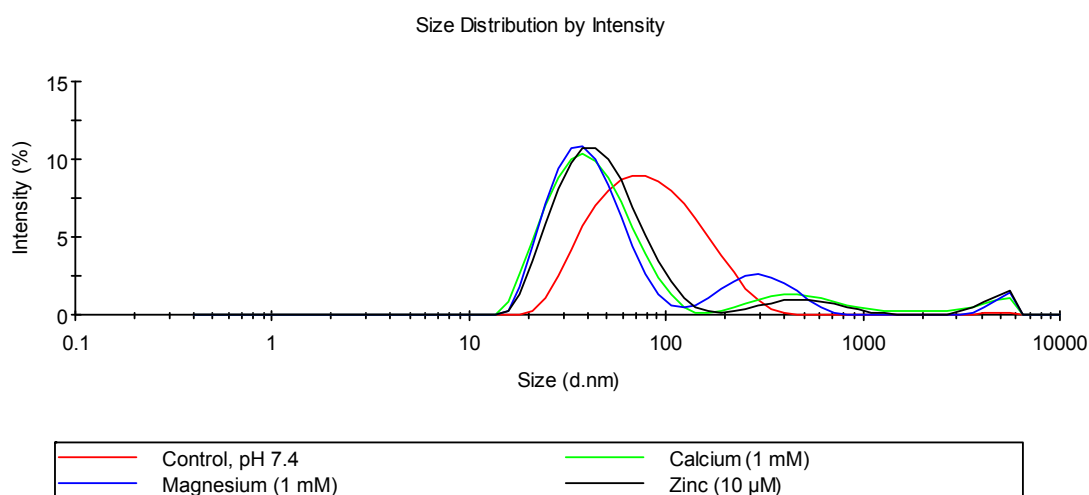


Figure 4.18: Size distribution of SEC-purified C1 in the presence of various metal ions at pH 7.4. The control sample shows a broad peak in the size distribution, showing that it contains a wide distribution of particle sizes with a Z-average of around 70 nm. In the presence of physiological concentrations of calcium, magnesium and zinc, the particle distribution is multimodal with three distinct populations.

In the absence of any metal ions, the sample contains a wide distribution of particle sizes, with a Z-average of around 70 nm. When physiological concentrations of calcium, magnesium and zinc are added to the C1 sample, the particle distribution becomes

multimodal, revealing three distinct populations. In the presence of magnesium, aggregates of around 300nm are observed, in addition to the more numerous 40nm particles (which are presumably the native C1 complex). The addition of calcium and zinc, on the other hand, results in the formation of 500nm particles.

Additive	Z-Average (d.nm)	Pdl	Pk 1 Mean Int (d.nm)	Pk 2 Mean Int (d.nm)
None	69.91	0.239	93.99	-
Magnesium	47.73	0.388	42.08	310.9
Calcium	44.57	0.363	44.44	540.6
Zinc	48.17	0.322	49.44	525.0

Table 14: The effect of calcium, magnesium and zinc on SEC-purified C1. When physiological levels of various metal ions are added to the C1 complex purified by SEC, two distinct populations are observed in the intensity distribution. The smaller particles appear to be around 40–50nm in diameter, while the size of the larger particles appears to vary with the nature of the metal ion. In the presence of magnesium, particles of approximately 300nm are seen, while the addition of calcium or zinc gives 500nm particles.

These results appear to indicate that the size distribution of SEC-purified C1 is affected in different ways by calcium, magnesium and zinc at physiological pH. However, this may not necessarily be the case. Since the majority of the particles in the sample correspond to C1 (with a mean size by number of around 40nm) it is possible that the larger particles observed in the intensity distribution are simply aggregating contaminants.

b) Calcium and copper at pH5.5

At sites of infection or inflammation, where C1 is likely to be found *in vivo*, the concentration of copper is thought to increase, accompanied by a decrease in pH of the local environment. Since this may have some bearing on the structure and function of C1, purified C1 was investigated by DLS in the presence of physiological concentrations of copper and calcium at pH5.5. The results of these experiments are illustrated below in Figure 4.19.

The size distribution profile of SEC-purified C1 in the presence of calcium remains similar to that observed at pH7.4. However, in the presence of copper, very large aggregates were observed. Copper forms an insoluble hydroxide in aqueous solutions which limits free Cu^{2+} concentrations to the order of 100 μM , but hydroxide formation is slow and can require hours to equilibrate (Hidmi and Edwards, 1999). As a further measure to prevent contamination with insoluble particles, copper solutions were filtered before being added to C1.

As with the previous studies on calcium, magnesium and zinc, the possibility that these are merely aggregates of contaminating proteins cannot be ignored. However, the possibility that the conditions of increased copper concentration and low pH induce aggregation of C1 is intriguing, and worth further investigation.

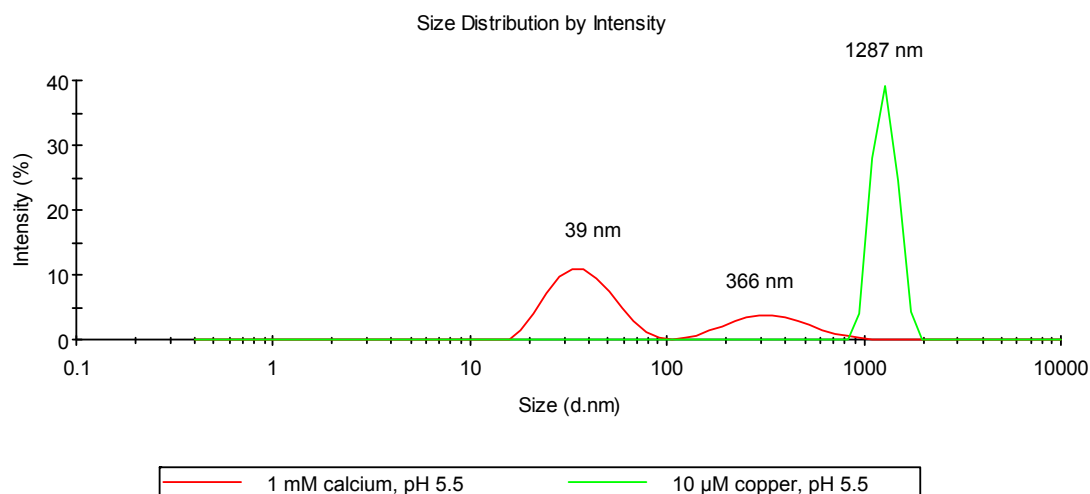


Figure 4.19: Size distribution of SEC-purified C1 in the presence of calcium and copper at pH5.5. At acidic pH, in the presence of physiological concentrations of calcium ions, the purified C1 sample shows a bimodal distribution of particle sizes, with peaks at 39nm and 366nm. In the presence of 10 μm copper chloride or copper sulphate, large aggregates were observed.

4.5.4 Melting point of SEC-purified C1

Finally, the effect of temperature on C1 purified by SEC was investigated by DLS in order to determine the melting point of the protein complex. This information may be useful in designing a crystallisation strategy, where temperature may be included as a variable. The results of these investigations are shown in Figure 4.20, and the melting temperature of the C1 complex is estimated to be around 40°C.

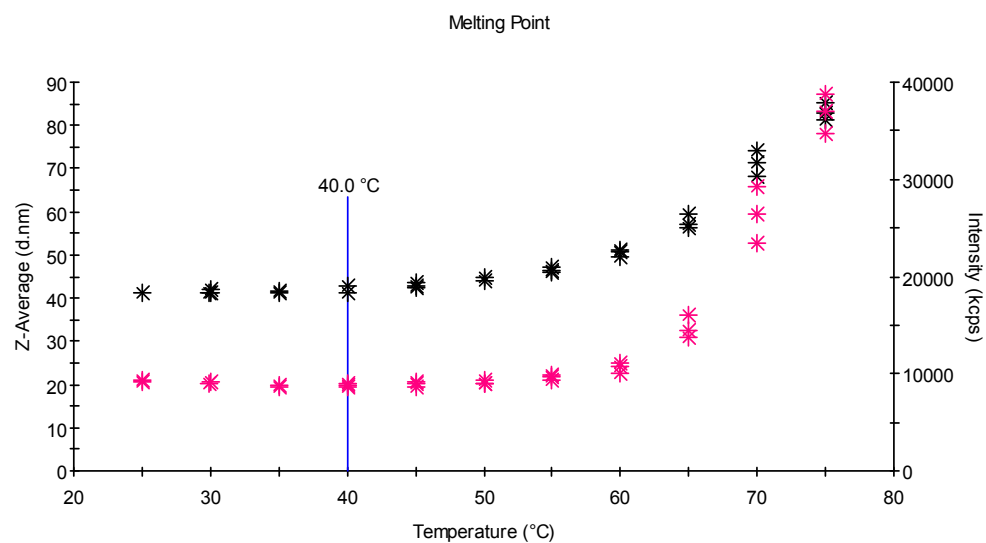


Figure 4.20: Melting point determination of SEC-purified C1. A sample of the pooled C1 peak was heated slowly to 75°C and three individual DLS measurements were recorded at 5°C intervals. Z-average (black) and intensity (magenta) are shown against temperature, and the melting point as determined by the Zetasizer Nano software is indicated by the blue line. At temperatures below 40°C the Z-average and intensity remain constant, while above 40°C they increase exponentially, indicating the presence of thermally denatured aggregates.

4.6 Crystallisation of C1 purified by SEC

Finally, complement component C1 that had been purified by the two step method outlined in section 4.3.3 was subjected to numerous crystallisation trials (see p. 130 for method). Initial attempts at crystallising the SEC-purified C1 complex began immediately the results of crystallisation attempts using partially purified C1 are described below.

4.6.1 Pre-crystallisation test

A pre-crystallisation test (PCT™) is available (Hampton Research, USA), which gives a useful indication of which protein concentrations are suitable for crystal screening purposes. Essentially, the test contains two reagents, one with a salt as the primary precipitant and one with a PEG as the primary precipitant. Furthermore, the two reagents are formulated at both a high and a low concentration, giving a total of four conditions. By examining the behaviour of the target protein under these conditions, protein concentrations which could produce an overabundance of unfavourable crystallisation conditions (such as clear drops or heavy precipitate) can be avoided. The PCT™ was used to determine the best concentration range for the SEC-purified C1 complex to be used in crystallisation trials.

The results of the PCT™ indicate that only the lowest concentration of C1 tested, 250 µg/ml, is immediately suitable for crystallisation trials. The results also show that crystallisation is unlikely to occur at concentrations approaching 10mg/ml. The results of the intermediate concentrations are ambiguous. In this case, it is recommended that the sample be tested to determine purity and homogeneity.

While C1 purified by this method appears to be relatively pure by native and SDS-PAGE (see section 4.4.1), the DLS data (see section 4.4.3) indicate that the

sample is heterogeneous and this is confirmed by the results of MS analysis, which show the presence of various contaminants (see section 4.4.2). Nonetheless, many proteins are known to crystallise from relatively crude preparations: a famous example being aldolase, which readily crystallises from homogenised rabbit muscle preparations (TAYLOR et al., 1948). Furthermore, when examining only those PCT™ conditions containing ammonium sulphate as the primary precipitant, it appears that protein concentrations between 0.25–1 mg/ml are entirely suitable for crystallisation. By contrast, C1 purified by size exclusion chromatography precipitated in all conditions containing PEG 4000 as the primary precipitant. This is hardly surprising given that the purification procedure exploits the ability of low concentrations of PEG 4000 to precipitate the protein from human serum.

Protein concentration (mg/ml)	A1	A2	B1	B2	Result
10	Amorphous	Amorphous	Granular	Amorphous	<i>Dilute sample</i>
1	Granular	Amorphous	Clear	Amorphous	<i>Diagnostic testing</i>
0.5	Granular	Amorphous	Clear	Amorphous	<i>Diagnostic testing</i>
0.25	Clear	Amorphous	Clear	Granular	<i>Perform screen</i>

Table 15: Results of the pre-crystallisation test on SEC-purified C1 at various protein concentrations. All conditions contain 0.1M Tris hydrochloride pH8.5. A1: 2.0M Ammonium sulphate; A2: 0.2M Magnesium chloride hexahydrate, 30% w/v Polyethylene glycol 4,000; B1: 1.0M Ammonium sulphate; B2: 0.2M Magnesium chloride hexahydrate, 15% w/v Polyethylene glycol 4,000.

The different effects of salts and PEGs on C1 can be explained by their different effects on protein stability and solubility (see section 2.4.1.1). Based on the results of the PCT™, one can predict that the C1 complex may be more responsive to crystallisation conditions where a salt is used as the primary precipitant, as opposed to PEGs, since commercially available crystallisation screens are typically formulated with high PEG concentrations. However, it is still unclear how different molecular weight PEGs or other organic polymers will affect the solubility of C1, or how it will behave using salts as the primary precipitants. In light of the findings outlined above, initial crystallisation screening with SEC-purified C1 was carried out at three protein concentrations — ranging between 0.3 and 1 mg/ml — using a variety of commercial and home-made screens. The results of these investigations are described in the next section and additional information can be found in Appendix I, section iii.

4.6.2 Crystallisation kit for protein-protein complexes

The Crystallisation Kit for Protein-Protein Complexes (Sigma-Aldrich Company Ltd., UK) is a sparse-matrix screen that was developed to reflect the most probable conditions for the crystallisation of protein complexes, based on a survey of the Protein Data Bank and the BMCD (Radaev et al., 2006). Accordingly, the screen heavily favours PEGs as the primary precipitants — as opposed to salts — since the results of the survey suggest that these conditions are more amenable to the crystallisation of protein complexes. Of the 48 conditions in the screen, approximately 80% contain PEG as the main precipitant. The results of the initial screens are shown in Table 16.

As predicted by the PCTTM, the results of the initial screen show that the C1 complex is much more soluble in conditions containing a salt as the primary precipitant. Even at the maximum protein concentration tested, precipitate was only observed in two thirds of the salt-containing conditions. In contrast, at protein concentrations above 0.3mg/ml, C1 precipitated in all conditions that contained PEG as the primary precipitant. The results also appear to indicate that a protein concentration of 0.3mg/ml is not sufficient to produce an appropriate degree of supersaturation in the crystallisation drop, since not one of the conditions yielded precipitated protein.

Concentration (mg/ml)	Salt conditions	PEG conditions
0.3	0/9	0/39
0.5	1/9	39/39
1	6/9	39/39

Table 16: The number of conditions in which SEC-purified C1 was observed to precipitate at various proteins concentrations. The primary precipitant is indicated as either salt or PEG.

The observed differences between using salts and PEGs as the primary precipitant can be explained by their distinct effects on protein stability and solubility, discussed in section 2.4.1.1. Given the observed differential behaviour of the C1 complex according to precipitant class, it would be reasonable to screen these two different classes of precipitants using two different protein concentration ranges. A concentration of 0.5 – 10mg/ml appears to be suitable for crystallisation conditions containing high salt concentrations, while 0.3–0.5mg/ml appears to be suitable for crystallisation where PEGs or other organic polymers are the main precipitants.

4.6.3 Stura Footprint Screen

The Stura Footprint Screen provides an alternative approach to screening crystallisation space using sparse matrix screens, which typically consist of empirically selected conditions that have been previously successful in crystallising various known proteins. Instead, the Stura Footprint Screen relies on testing the relative protein solubility against various common precipitants in a systematic fashion. Again, the screen is based on previously successful crystallisation conditions, which have been expanded into a format that allows systematic screening of the protein–precipitant solubility curve. The screen mainly contains PEGs as the primary precipitant. While the conditions in the Stura Footprint Screen are also empirically determined, the results of the screen may present a more detailed picture of the solubility of the protein of interest under a range of conditions.

SEC-purified C1 was tested at a concentration of 0.4 mg/ml. This concentration was chosen based on the results of the Crystallisation Kit for Protein-Protein Complexes, which indicates that a C1 concentration between 0.3–0.5 mg/ml is likely to precipitate an intermediate number of conditions when PEG is used as the primary precipitant. After one month, 4/48 conditions had precipitated, while the rest remained clear, and no further changes were observed thereafter. This indicates that the concentration of protein used may be too low and that successful crystallisation of C1 may require protein concentrations in excess of 0.4 mg/ml.

4.6.4 Structure Screen 1 & 2

Structure Screen 1 & 2 (Molecular Dimensions, UK) are classic sparse matrix screens, which include successful crystallisation conditions from previously published works. The screens were designed for typical globular proteins, rather than protein complexes, and thus feature a greater diversity of conditions containing salts as the major precipitant than the previously tested Crystallisation Kit for Protein-Protein Complexes.

C1 was tested at a slightly increased concentration of 0.5 mg/ml, and numerous conditions produced large, three-dimensional crystals. However, on testing using the various methods described in section 2.4.2.3, these were largely revealed to be salt crystals. Nevertheless, several other interesting conditions could be identified and these are shown in Table 17.

Reference	Precipitants	Buffer	Additive	Result
SS1–27	1.5M lithium sulphate	0.1 M HEPES pH7.5	n/a	Phase separation
SS1–35	30 % w/v PEG 4000 0.2M lithium sulphate	0.1 M Tris pH8.5	0.2M lithium sulphate	Phase separation
SS2–6	1 M lithium sulphate	0.1 M Tris pH8.5	0.01 M nickel chloride	Phase separation
SS2 -27	25 % v/v PEG MME 550	0.1 M MES pH6.5	0.01 M zinc sulphate	Crystals
SS2–38	30 % v/v PEG 400	0.1 M sodium acetate pH4.6	0.1 M cadmium chloride	Phase separation
SS2–39	30 % w/v PEG MME 2000	0.1 M sodium acetate pH4.6	0.2M ammonium sulphate	Phase separation

Table 17: Table of hits from Structure Screens 1 & 2 using SEC-purified C1. These conditions were identified as being possible conditions for optimisation of C1 crystallisation using variables such as protein concentration, precipitant concentration and pH. The primary precipitant, buffer and any additives for each hit condition are indicated.

Phase separation is often an indication that the conditions within the crystallisation experiment are close to those required for crystallisation. In contrast to the other crystals observed in these screens, those obtained using 25% PEG MME 550 in 0.1 M MES pH6.5 and 0.01 M zinc sulphate were easily destroyed by mechanical stress, indicating that these may indeed be protein crystals. This was further supported by the observation that the crystals appear to be growing from the surrounding protein precipitate, with clearly visible halos of depleted precipitate around the nascent crystals, as shown in Figure 4.21.

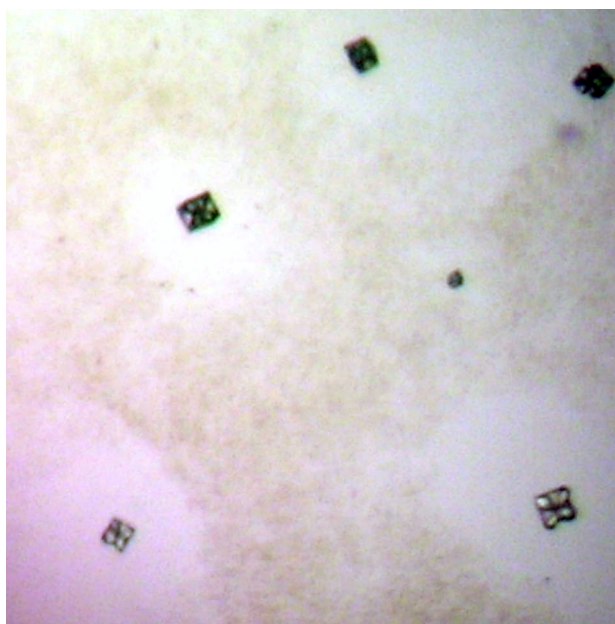


Figure 4.21: Crystals apparently growing from precipitated protein in the crystallisation drop. The crystallisation conditions were 25% PEG MME 550 in 0.1 M MES pH6.5 and 0.01 M zinc sulphate. Two morphologies can be distinguished: cubic crystals with interior faces (top) and irregular quatrefoils (bottom). Despite appearing to grow from precipitated protein, the crystals, if they diffracted at all, showed diffraction patterns characteristic of salt crystals.

A number of crystals were tested for diffraction. The mother liquor vitrified on flash freezing, meaning that no additional cryoprotectant was necessary. As can be seen from

Figure 4.21, the crystals appeared in two distinct morphologies: cubic crystals with interior faces; and irregular quatrefoils. After numerous sacrifices, one quatrefoil was successfully split into its four separate lobes. Each lobe was mounted individually, and three of the cubic crystals were also tested. Unfortunately, the crystals either did not diffract at all, or at best produced a few strong, widely spaced reflections, which is characteristic of the diffraction pattern of salts. The depleted salt concentration in the vicinity of the crystals may increase the solubility of the protein in this region, explaining the halos.

Optimisation of the crystallisation conditions in terms of precipitant concentration and protein concentration did not produce any improvement in crystal diffraction. Screening around the other interesting conditions identified in the screens also failed to produce any further protein crystals. Since the C1 complex purified by SEC alone still contains various contaminating proteins, it may be that these are preventing C1 from crystallising. So, the next step in the search for crystals of native complement component C1 was to attempt to further purify the complex to remove any remaining contaminants.

4.7 Further purification by ion exchange chromatography

All attempts to crystallise C1 purified by size exclusion chromatography using the optimised procedure (see section 4.3.3) were unsuccessful. In a bid to improve the chances of crystallisation by removing the remaining contaminants, C1 was further purified by ion exchange chromatography. As previously described, IEX separates proteins according to their net surface charge. Charged proteins bind reversibly to the oppositely charged IEX medium: a protein with $pI < pH$ will bind to a anion exchanger, while when $pI > pH$, the protein will behind to a negatively charged cation exchanger. For C1, the predicted pI is 7.2, so under physiological pH (7.4) the complex is predicted to have a slight negative charge and will thus bind to an anion exchanger such as Mono QTM.

Figure 4.22 shows the separation of the SEC-purified C1 into three major fractions by IEX. The first peak contains unbound proteins, which appears to be various contaminants, while intact C1 appears to elute in peak B. The corresponding SDS-PAGE gel shows two clear bands at around 75kDa (C1r and C1s) as well as a faint triplet at around 25kDa (C1q). There are also two weaker bands at around 50 kDa, which may be traces of the contaminants that did not bind to the column. Alternatively these could be the heavy fragments of cleaved C1r and C1s (51 and 47 kDa, respectively), with the light fragments (27 and 28kDa) being almost indistinguishable from the various C1q chains. The third peak shows a similar pattern of bands to the C1 peak, although the bands differ in intensity, with the band corresponding to C1s appearing more prominent and C1r and the C1q triplet appearing much weaker.

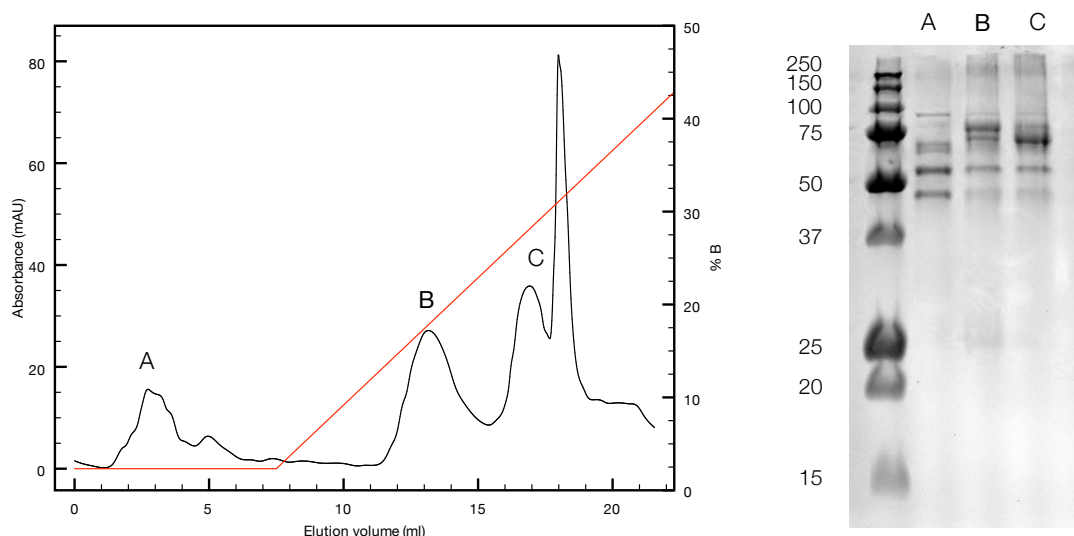


Figure 4.22: Further purification of C1 using a Mono Q IEX column. The starting buffer (buffer A) is the same as that used for the elution of C1 in the previous chromatography step, while a high-salt buffer (buffer B) is used to elute bound proteins from the column. An SDS-PAGE gel corresponding to the three peaks collected is also shown. Buffer A: 10mM Tris pH7.4, 150mM sodium chloride; buffer B: 10mM Tris pH7.4, 1M sodium chloride. A. Unbound proteins. B. Intact C1 complex. C. Unknown.

In the chromatogram, the two peaks are not fully resolved, so a likely explanation for the observed results is that peak C contains a small amount of the tail end of peak B, while predominantly containing an unrelated ~ 75 kDa protein. An alternative explanation is that the second peak contains a glycosylated form of C1. Glycosylation can affect the staining of proteins by Coomassie dyes (Darren Thompson, personal communication), so different degrees and types of glycosylation of the C1 complex could explain the observed differences in the intensity of the bands between the two fractions. Alternatively, the third peak may contain some combination of C1r and C1s, for example the C1r₂C1s₂ tetramer, which may exist independently of C1q (Arlaud et al., 1980).

It has previously been shown that the C1 complex dissociates when purified by ion exchange chromatography in the presence of 10mM EDTA (Lepow et al., 1963). In an attempt to purify the complex further, the Mono Q separation was performed again

at pH8 using 10mM EDTA in the loading and elution buffers, and the results are shown in Figure 4.23. In addition, the gradient was modified in an attempt to get a better separation of the two peaks of interest. As can be seen from the gel, the separation of C1 from the 50kDa contaminants is much cleaner: the ~50kDa bands are found exclusively in the unbound fraction, for example.

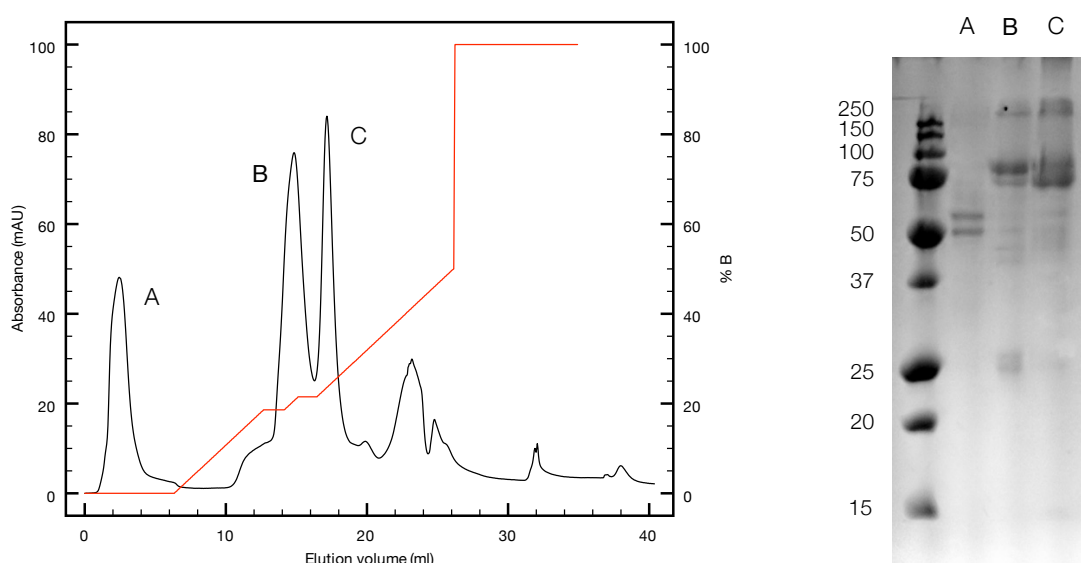


Figure 4.23: A modified further purification of C1 using a Mono Q IEX column. In this experiment, the pH was increased and the gradient was held as each peak began to elute. Again the SDS-PAGE gel corresponding to each peak is shown. Buffer A: 10mM Tris, pH8, 150mM sodium chloride, 10mM EDTA; buffer B: 10mM Tris, pH8, 1M sodium chloride, 10mM EDTA. A. Unbound proteins. B. Intact complement component C1. C. Unknown. The results show that the separation is dramatically improved using this modified method, as the eluted peaks contain fewer contaminating proteins.

The improved separation appears to be a consequence of modifying the pH and the shape of the elution gradient, rather than the inclusion of EDTA in the loading and elution buffers. In fact, the presence of EDTA does not appear to disassemble the complex, as expected. This observation is further supported by the results of dynamic

light scattering studies, as described in section 4.4.3 and section 4.9.3. Further purifications involved the use of longer step-gradients to give optimal separation of the individual protein peaks (data not shown).

4.8 Summary: Purification by SEC and IEX

After a first purification step using SEC (see p. 125 and 157), fractions containing C1 are pooled and exchanged into 10mM Tris, pH8, 150mM NaCl, before being applied to a Mono QTM 5/50 GL column equilibrated with the same buffer. Unbound proteins are washed away and the bound proteins are eluted with a step-gradient of buffer B (10mM Tris, pH8, 500mM NaCl). Two major peaks elute from the column, at 18% and 23% concentration of buffer B. When the corresponding fractions are analysed by SDS-PAGE, the first peak appears to contain all components of the C1 complex, while the second peak appears to contain two proteins of approximately 75kDa. These may represent a contaminant, a differently glycosylated form of C1, or possibly the C1_r₂C1_s₂ tetramer.

4.9 Characterisation of Mono Q-purified C1

In order to investigate the results of separating SEC-purified C1 by anion exchange using Mono Q, the two eluting peaks were subjected to analysis by a number of complementary methods. Since the aim of this project is to obtain native C1 in a form suitable for use in crystallisation trials, the techniques employed were chosen to determine as clearly as possible whether the purified product obtained is indeed C1, and whether it is pure and structurally homogeneous. This following section describes the results of these investigations, and indicates that the quality of purified C1 has been improved by using IEX as a second chromatography step in the purification procedure.

4.9.1 SDS-PAGE analysis of Mono Q-purified C1

Since native agarose gel electrophoresis does not yield any additional useful information to that obtained by SEC, the characterisation of Mono Q-purified C1 by electrophoretic methods was restricted to SDS-PAGE. In addition to being able to resolve individual polypeptide chains in a sample, the separation can be performed with or without reducing agent, which can give information about the presence of any disulphide bonds. The detergent in the sample buffer causes proteins to unfold into linear polypeptide chains. However, in the absence of reducing agents, any chains linked by disulphide bonds will remain associated. Native C1 contains numerous disulphide bonds within and between the individual components of the complex. Since investigations of reduced and oxidised C1 that had been purified only by SEC yielded unusual results (see section 4.4.1.2), analysis of the Mono Q-purified C1 was also carried out in the presence and absence of the reducing agent BME. The results are shown in Figure 4.24.

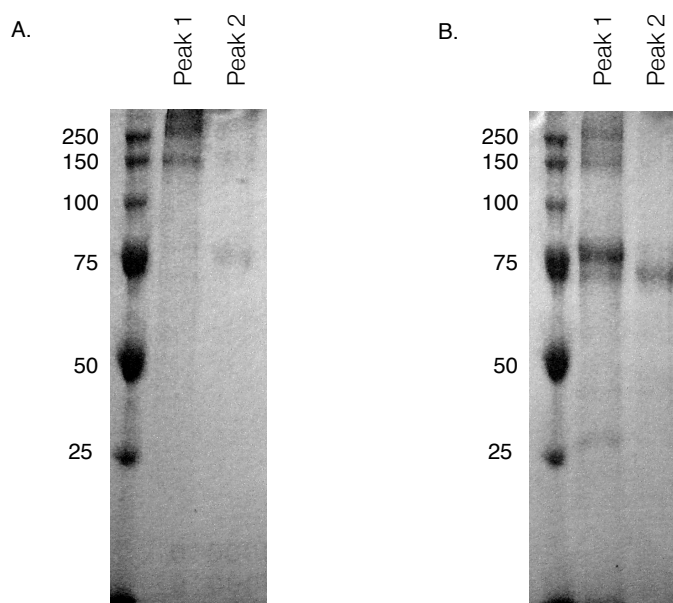


Figure 4.24: Oxidised and reduced SDS-PAGE gels of the two major peaks eluting from a MonoQ separation of SEC-purified C1. A. Samples run in the absence of a reducing agent show mainly high-molecular-weight-aggregates, some of which did not fully enter the gel, as well as a faint band at around 75kDa in Peak 2. B. The same samples run in the presence of BME show that Peak 1 contains all the C1 sub-components (C1r and C1s at ~75kDa; C1q at ~25kDa), as well as some high-molecular-weight aggregates, while Peak 2 contains a protein of approximately 75kDa.

The results show that C1 further purified by IEX behaves comparably to C1 simply purified by SEC alone when subjected to SDS-PAGE in the absence of reducing agent. Two bands at approximately 150 and 250kDa are observed, which are also seen with reduced intensity under reducing conditions. Interestingly, these are close to the predicted molecular weights of dimers composed of C1r and/or C1s (~150kDa) and the C1r₂C1s₂ tetramer (~300kDa). The bands may simply be contaminants, or they may represent disulphide-linked forms of the C1r and C1s components.

In addition, the results show that there is unlikely to be any intact C1 at all in the second peak eluting from the Mono Q column. A single ~75kDa band is seen under reducing conditions and is also faintly visible when the sample remains oxidised. This indicates that the unknown protein is linked by disulphide bonds to form a complex that is too large to enter the gel.

4.9.2 Peptide mass fingerprinting of MonoQ-purified C1

Solution samples of the two major peaks eluting from further purification of C1 by anion exchange were submitted for peptide mass fingerprinting (see page 127). The aim of these investigations was to determine the nature of the proteins eluting from the Mono Q anion exchange column during the final purification stage. Since each peak is likely to contain only a few proteins — and in order to minimise handling and exposing the samples to sources of contamination — peptide mass fingerprinting and MS/MS analysis were carried out directly on the solution samples.

The samples contain most of the same contaminating proteins identified previously in SEC-purified C1 (keratins, immunoglobulins and other abundant serum proteins). However, of 44 proteins identified with a minimum probability of 99.99%, 15 were complement-related proteins. C1r and C1s were conclusively identified in both major peaks eluting during the final purification, while C1q was only detected in the first peak. Complement-related proteins identified in Mono Q-purified C1 and their distribution between the two peaks are shown in Table 18 and Figure 4.25.

Table 18: Complement-related proteins identified in Mono Q-purified C1. The table shows the complement-related proteins identified in each peak eluting from the final purification step. The molecular weight, protein identification probability and peptide sequence coverage is indicated for each protein identified in the sample. C1r and C1s are present in both peaks, while C1q was detected in the first peak only.

	MW (kDa)	Peak 1		Peak 2	
		Probability (%)	Sequence coverage	Probability (%)	Sequence coverage
Apolipoprotein B-100	516	100		100	
C4-A	193	100		100	
C3	187	100		100	
Factor H	139	100		0	N/A
C1r	82	100		100	
MASP-1	82	N/A	N/A	50	
C1s	77	100		100	
Vitamin K-dependent protein S	75	100		100	
C4BP (α chain)	67	100		100	
Galectin-3 binding protein	65	N/A	N/A	100	
C9	63	N/A	N/A	50	
C1-INH	55	100		100	
C4BP (β chain)	28	100		100	
C1q	26	55		0	N/A

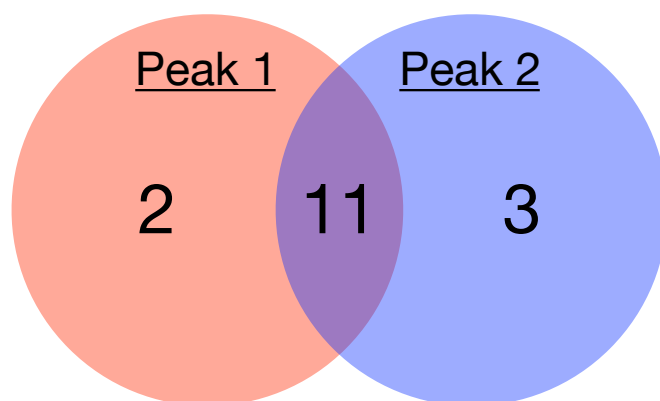


Figure 4.25: Venn diagram showing the distribution of unique, complement-related serum proteins between the two major peaks eluting from Mono Q purification of complement component C1. Minimum protein probability: 50 %; minimum peptide probability: 95 %. Proteins detected in peak 1 but not in peak 2 were: complement factor H and C1q. Proteins in peak 2 but not peak 1 were: galectin 3-binding protein, complement component C9 and mannan-binding lectin serine protease 1 (MASP-1).

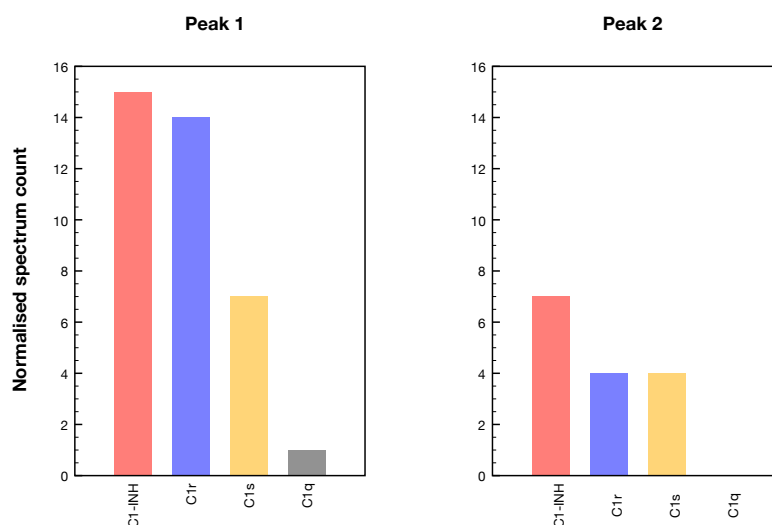


Figure 4.26: Quantitative analysis of C1 subcomponents and C1 inhibitor identified in the two major peaks eluting from Mono Q purification of C1. The normalised spectrum count provides quantitative information about relative abundances between samples. It is calculated by averaging the number of identified spectra in each sample and then the number of spectra assigned to a given protein is multiplied by the ratio of the average spectrum count to the number of spectra in that sample. The first peak to elute from the Mono Q column gives a high normalised spectrum count for C1-INH and C1r, with C1s and C1q detectable to a lesser extent. In the second peak to elute, the normalised spectrum count of C1-INH and the various C1 subcomponents is decreased, with C1q not detectable at all. This suggests that the first peak contains the native C1 complex, while the second peak may contain smaller amounts of C1 (or the C1₂C1s₂ tetramer).

When the normalised spectrum count of C1 subcomponents in each peak is compared, we see C1 appears to be enriched in the first peak, with C1q also present in the sample (Figure 4.26). Apolipoprotein B-100 and complement component C3 appear to have equal normalised spectrum counts across both samples, while the remaining proteins are more abundant in the second peak (data not shown). These results, in combination with the information from SDS-PAGE analyses, indicate the first peak contains primarily the intact native C1 complex.

In the second peak the normalised spectrum count of the various C1 subcomponents is decreased, while C1q was not detected in the sample at all. This suggests that the second peak may contain smaller amounts of C1 (or the C1r₂C1s₂ tetramer). However, C1q is known to be resistant to trypsinolysis (Reid, 1979; Tissot et al., 2005), which may also explain the low abundance of C1q peptides. C1-inhibitor (C1-INH) was also identified in both samples, with a similar normalised spectrum count profile to C1r and C1s.

Two proteins could be identified with a minimum protein probability of 50% in the first sample, but not the second. These were complement factor H and C1q. Conversely, three proteins were detected in the second sample but not the first: galectin 3-binding protein, complement component C9 and mannan-binding lectin serine protease 1 (MASP-1). The results suggest that the native C1 complex is enriched in the first peak eluting from the final purification step, while the second peak contains mainly contaminants.

4.9.3 Dynamic light scattering of Mono Q-purified C1

The results outlined in section 4.4.3 show C1 purified by size exclusion chromatography is polydisperse and shows two broad peaks in the size distribution by intensity. The first peak is centred at approximately 45nm, which is close to the proposed diameter (40nm) of the assembled C1 complex (Perkins, 1985). The second peak is at around 180nm. The majority of particles in the sample are the smaller, 45nm particles (for an overview of interpreting DLS data, see section 2.3.3.2). Finally, the width of the distribution indicates that there is likely to be a range of particle sizes, likely due to the presence of various different contaminants that could be removed by further purification using ion exchange chromatography. The various fractions eluting from the Mono Q column were analysed by DLS, and the results are shown in Table 19.

Sample	Pdl	Z-average (d.nm)	MW (kDa)
Unbound	0.317	17.63	548
Peak 1	0.292	27.16	1510
Peak 2	0.124	27.44	1540

Table 19: Dynamic light scattering analysis of the various fractions obtained by further purification of complement component C1 by IEX.

The second peak eluting from the column is likely to contain the intact C1 complex, as determined by SDS-PAGE and peptide mass fingerprinting. While the sample and buffers were all filtered through a 0.2 μm filter during further purification by the Mono Q method, purified C1 contains particles of around 200 nm that can be removed by further filtration. The results of dynamic light scattering studies on purified C1 are shown in Figure 4.27 and Table 20.

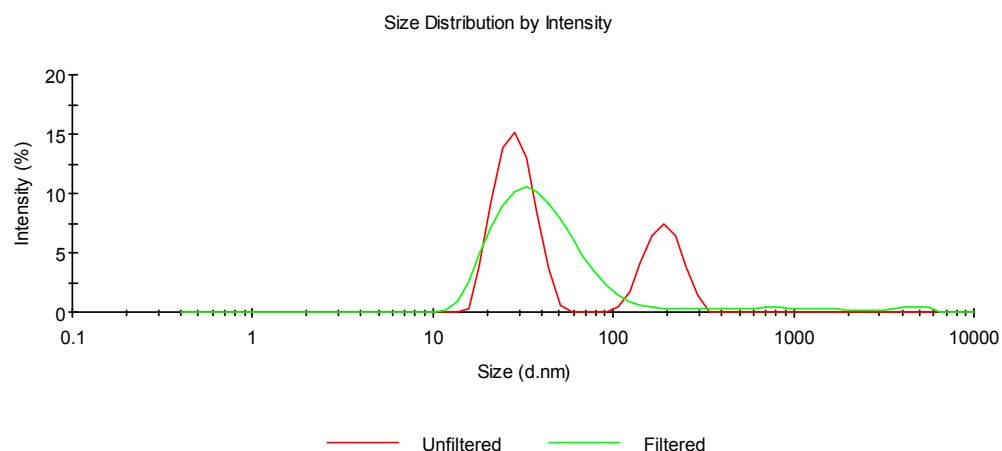


Figure 4.27: Size distribution by intensity of Mono Q-purified C1. The unfiltered sample shows a bimodal distribution, while the filtered sample shows a monomodal distribution, indicating that the larger particles may be removed by filtering through a 0.2 μm filter.

The intensity plot of the unfiltered sample shows a bimodal size distribution, with the majority of particles around 30 nm in diameter and a small population of larger 200 nm particles, according to multimodal analysis. When the sample is filtered, the size distribution becomes monomodal, with a Z-average size of 36 nm.

Sample	PdI	PdI Width (d.nm)	Z-Average (d.nm)	Pk 1 mean int. (d.nm)	Pk 2 mean int. (d.nm)
Unfiltered	0.252	59.55	118.6	28.94	192.4
Filtered	0.278	19.12	36.25	43.34	-

Table 20: DLS analysis of Mono Q-purified C1. The protein concentration was 0.1 mg/ml in 10 mM Tris, pH 7.4, 150 mM NaCl. The unfiltered sample shows a bimodal size distribution, with the majority of particles around 30 nm in diameter. By filtering the sample through a 0.2 μm filter, the size distribution becomes monomodal. The results of cumulants and multimodal analysis correspond, giving a particle size of around 40 nm for Mono Q-purified C1.

The results of cumulants and multimodal analysis are in close agreement (36.25 nm and 43.34 nm respectively) and the PDI width substantially decreases when the sample is filtered, indicating that a single population of molecules is present in the sample. However, the polydispersity remains high, which suggests that while the particles are of relatively uniform size, some structural diversity still exists within the population of purified C1 molecules.

4.9.4 Storage of Mono Q-purified C1

To test the effects of storing C1 for a prolonged period of time, a freshly purified sample was analysed by DLS and stored at 4°C for 30 days. The sample was then re-analysed and the results are shown in Table 21. It appears that C1 aggregates slightly upon storage under the conditions tested. However, that the high-molecular-weight material could be removed by filtration through a 0.2 µm filter. The resulting protein solution was comparable to freshly purified C1 in terms of particle size and polydispersity. Therefore C1 further purified by IEX can be stored at 4°C for the period of one month at least without significant degradation in sample quality, and any aggregates can be simply removed.

Day	Z-Average (d.nm)	Pdl
0	29.86	0.205
30	33.38	0.268
Difference	+ 3.52	+ 0.063

Table 21: Effect of storing Mono Q-purified C1 at 4°C. The protein was stored in 10mM Tris, pH8, 150mM NaCl. After 30 days, a small increase in the Z-average size and the polydispersity was seen, indicating some aggregation had occurred. However, the aggregates could be removed by filtration to restore the original values.

4.10 Mono Q-purified C1 under relevant conditions

Purified C1 (see section 4.8 for purification details) was investigated under various conditions to determine the effects of ionic strength, pH, and biologically relevant divalent metal cations on the complex. The results of these experiments are described in this section. These investigations were designed to complement the experiments previously performed on SEC-purified C1 (see section 4.5), and formed the basis of projects completed by other members of our laboratory.

4.10.1 The effect of ionic strength on Mono Q-purified C1

Published reports suggest that the function of the C1 complex is affected by changes in the ionic strength of the surrounding medium due to dissociation of the complex at high ionic strength (Lepow et al., 1958; Colten et al., 1968). However, the investigations detailed in section 4.5.1 showed that ionic strength does not affect the solution structure of SEC-purified C1 as would be expected if C1 were dissociating at high ionic strength. Distinct size distribution profiles emerge depending on the ionic strength of the sample buffer.

When no additional salt is present in the buffer, SEC-purified C1 is polydisperse with a Z-average size of around 96nm. In the presence of physiological salt concentrations, the Z-average size decreases to around 80nm and the size distribution becomes monomodal. At high ionic strength the Z-average size increases slightly to around 86nm with no further change in polydispersity. These results suggest that fluctuations within the physiological range of ionic strength do not appreciably affect the structure of SEC-purified C1, while extreme values caused the sample to aggregate, rather than dissociate. It is not clear whether this effect was due to the presence of contaminants that were still present in the SEC-purified sample.

Further purification by IEX removed several remaining contaminating proteins, so Mono Q-purified C1 was analysed for comparison. The effect of various salt concentrations on the size distribution by DLS is shown in Table 22. Again, distinct profiles emerge, although the difference between them is smaller. In the physiological range, the sample has a Z-average size of around 40nm. At high ionic strength, this increases slightly to around 42nm. From multimodal analysis, we can see that the change in Z-average size corresponds to an increase in the mean size of the first peak in the intensity distribution.

NaCl (mM)	Z-Ave (d.nm)	Pdl	Pk 1 Mean Int (d.nm)
150	40.99	0.375	36.28
200	40.08	0.365	35.99
300	40.09	0.367	35.03
500	41.68	0.374	36.86
1000	41.54	0.360	39.10

Table 22: The effect of ionic strength on the size and distribution of Mono Q-purified C1 by DLS. Distinct states can be determined, as for SEC-purified C1, except that the difference between them is much subtler. In a low ionic strength buffer at pH7.4 and in the presence of physiological salt concentrations, the sample is polydisperse with a Z-average size of around 40nm. At high ionic strength, the Z-average increases to around 42nm.

These results suggest that high ionic strength does not in fact dissociate the C1 complex purified by the method outlined in section 4.8; a small increase in the size of the 30–40nm particles is observed instead. Since these are the most numerous particles in the sample, it is reasonable to conclude that the C1 complex is stable across the physiological range of ionic strengths at pH7.4, while high salt concentrations cause a slight increase in the hydrodynamic radius of the molecule. This may indicate a conformational change, which could explain the loss of function described by others.

4.10.2 The effect of EDTA on Mono Q-purified C1

The addition of EDTA is known to dissociate C1 into its individual sub-components (Lepow et al., 1963). Adding increasing concentrations of EDTA to SEC-purified C1 causes aggregation of the sample (see Figure 4.17, p. 183), although whether this is caused by aggregation of the C1 complex or contaminating proteins remains unclear. Since the majority of contaminating proteins could be removed by further purification of the complex, the Mono Q-purified C1 was analysed in the presence of various EDTA concentrations, and the results are presented in Figure 4.28, Figure 4.29 and Table 23.

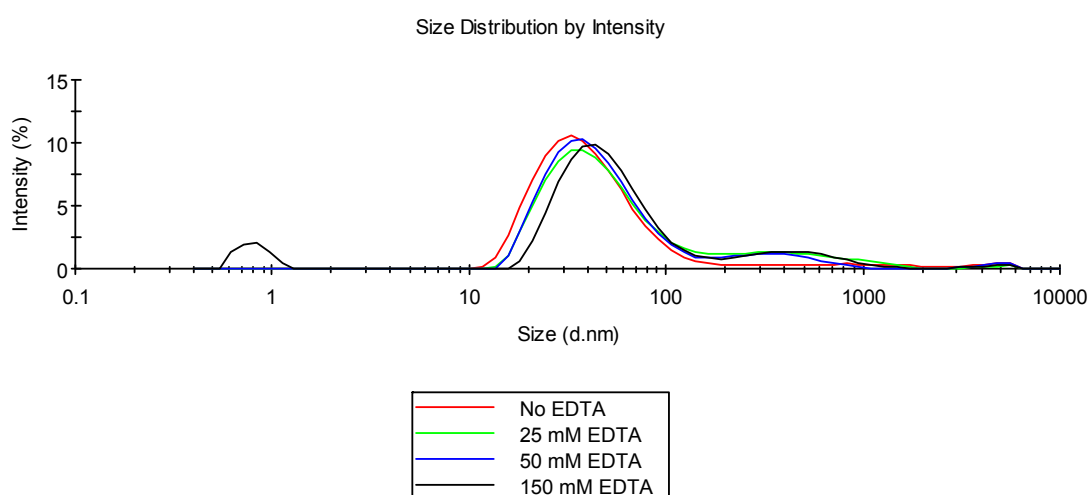


Figure 4.28: The effect of EDTA on the size distribution by intensity of Mono Q-purified C1.

As can be seen from the intensity distribution, adding EDTA to purified C1 causes a small amount of larger aggregates to form. At high EDTA concentrations (150mM), some very small particles are observed, in addition to a similar distribution of 40nm particles (which correspond to the proposed diameter of the native C1 complex). The 0.75nm particles observed in the presence of high concentrations of EDTA correspond to a predicted molecular weight of 1.7kDa. This is extremely small, and thus does not

suggest that C1 is dissociating into its individual subcomponents. These observations are consistent with the results presented in for SEC-purified C1, but are in stark contrast to published literature on the subject.

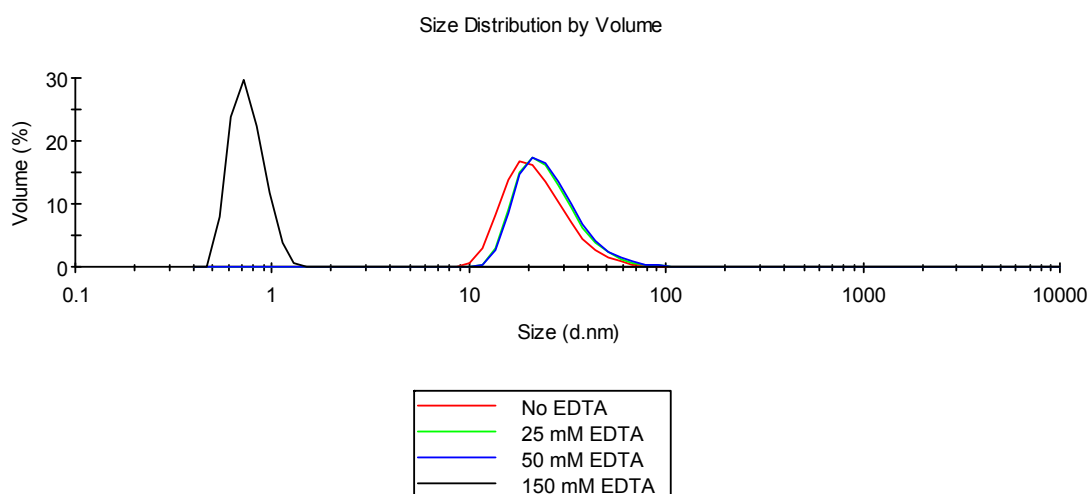


Figure 4.29: The effect of EDTA on the size distribution by volume of Mono Q-purified C1. When the intensity distribution is transformed into a volume distribution, we see that the 0.75nm particles make up 30% of the scattering matter in the sample at high EDTA concentrations. The increase in volume at low EDTA concentrations is also more obvious.

EDTA (mM)	PdI	Z-Average (d.nm)	Pk 1 intensity (d.nm)	Pk 2 intensity (d.nm)	Pk 1 volume (d.nm)	Pk 2 volume (d.nm)
0	0.278	36.25	43.34	886.2	23.56	-
25	0.358	42.95	50.47	534.6	26.60	-
50	0.335	39.90	47.06	378.1	26.87	-
150	0.431	43.34	53.64	500.4	0.7543	32.37

Table 23: The effect of EDTA on Mono Q-purified C1. The sizes determined by volume rely on a number of assumptions, and should be used as a guide only.

4.10.3 Calcium and EDTA at acidic and neutral pH

To complement the studies on SEC-purified C1 (see section 4.5.3), Mono Q-purified C1 was analysed by DLS at acidic and neutral pH. In addition, the effect of calcium and EDTA under these conditions was investigated. The results of these experiments are presented in this section.

4.10.3.1 Mono Q-purified C1 at physiological and acidic pH

When the Mono Q-purified sample was exchanged into a low pH buffer (10mM MES, pH 5.5), no significant changes were observed in the size distribution profile. The results of DLS analysis at physiological and acidic pH are shown in Table 24. Although the Z-average size appears to increase, in terms of volume the particles that constitute the majority of the scattering content of the sample remain virtually identical in size at both pH values. The function of the C1 complex is reported to be greatly affected by pH (Lepow et al., 1958), with a sharp decrease in activity below pH6.5. This may be due to conformational changes that render the complex inactive, since C1 does not appear to dissociate or aggregate significantly at pH5.5.

pH	Pdl	Z-Average (d.nm)	Pk 1 intensity (d.nm)	Pk 2 intensity (d.nm)	Pk 1 volume (d.nm)
5.5	0.448	50.33	44.13	260.6	27.30
7.4	0.377	45.52	42.41	332.2	27.95

Table 24: The size distribution profile of Mono Q-purified C1 at acidic and physiological pH. No significant changes are observed between pH7.4 and pH5.5.

4.10.3.2 The effect of calcium on Mono Q-purified C1

a) Calcium and EDTA at pH 7.4

When physiological concentrations of calcium are added to SEC-purified C1, the particle distribution is multimodal: the more numerous 40 nm particles correspond to the proposed diameter of native C1, while larger 500 nm particles may be aggregating contaminants. Since these aggregates were likely removed by the subsequent purification step, the effect of calcium on Mono Q-purified C1 was determined at physiological pH. After addition of 2 mM calcium chloride to the sample (which corresponds to physiological levels of calcium ions in human blood plasma), the sample was further titrated with EDTA. The results of these investigations are presented in Figure 4.30 and Table 25. Physiological concentrations of calcium do not appreciably affect the size of purified C1 determined from the intensity distribution (42.41 nm and 43.01 nm for the control and calcium samples, respectively). Adding EDTA to the sample produced results comparable to those detailed on page 217.

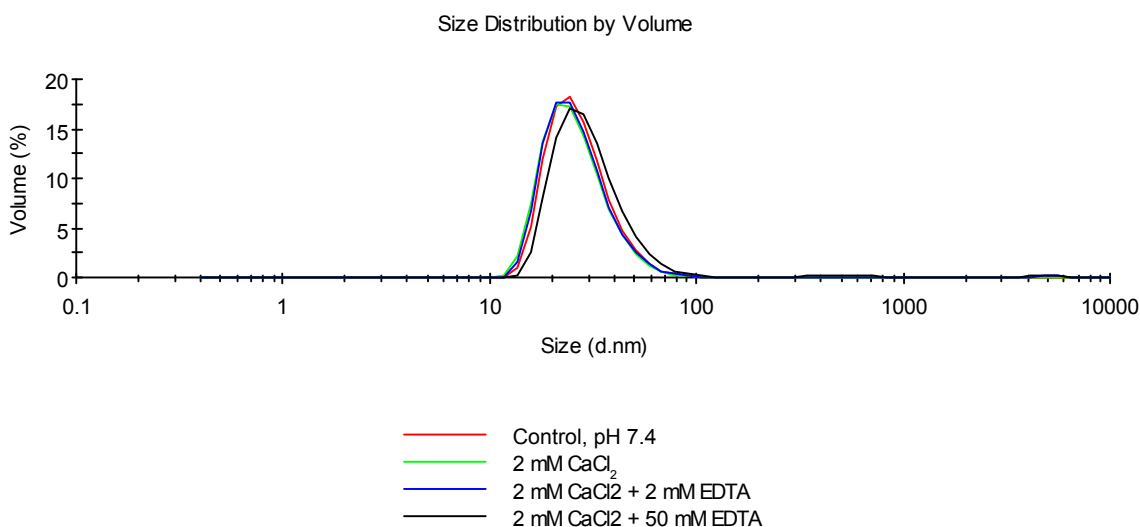


Figure 4.30: Effect of calcium and EDTA on the size distribution by volume of Mono Q-purified C1 at pH 7.4. No appreciable changes are observed in the presence of 2 mM calcium at physiological pH. The addition of EDTA causes a slight increase in the mean particle size.

Sample	Pdl	Z-Average (d.nm)	Pk 1 intensity (d.nm)	Pk 2 intensity (d.nm)	Pk 1 volume (d.nm)	Pk 2 volume (d.nm)
Control	0.377	45.52	42.41	332.2	27.95	424.1
2mM CaCl ₂	0.331	42.78	43.01	362.1	26.89	484.9
2mM CaCl ₂ + 2mM EDTA	0.419	53.50	43.89	305.2	27.32	389.8
2mM CaCl ₂ + 50mM EDTA	0.357	68.02	52.39	384.3	31.08	489.5

Table 25: The effect of calcium and EDTA on Mono Q-purified C1 at pH7.4. Adding physiological concentrations of calcium to the sample does not affect the size of the purified C1 complex, as can be seen from the volume distributions.

b) Calcium and EDTA at pH5.5

The local environment at sites of infection or inflammation — where C1 is likely to be found *in vivo* — is thought to be acidic, at around pH5.5. SEC-purified C1 does not appear to be greatly affected by physiological calcium concentrations at low pH, with the size distribution profile remaining similar to that observed at pH7.4. For comparison, the size distribution profile of Mono Q-purified C1 is shown in Figure 4.31 and Table 26. Physiological concentrations of calcium do not appreciably affect the size of purified C1 determined from the intensity distribution (44.13nm and 43.11nm for the control and calcium samples, respectively), and these results closely correspond to those obtained at pH7.4 (see page 220).

However, when the distribution is converted into a volume distribution, it appears that there is a subtle increase in particle size on the addition of calcium, which is not reversible by the subsequent addition of EDTA. In the absence of any additional calcium, the C1 complex appears to have a hydrodynamic radius of around 27nm. On the addition of 2mM calcium, this appears to increase to approximately 32nm, an increase of 5nm overall. While this is an interesting finding, transforming the original size distribution by intensity into a distribution by volume relies on a number of assumptions, and amplifies any errors in the original analysis. Therefore, the solution behaviour of purified C1 in the presence of various biologically relevant divalent metal cations is worth further investigation. However, this was deemed a suitable point to conclude these particular experiments within the context of the present work.

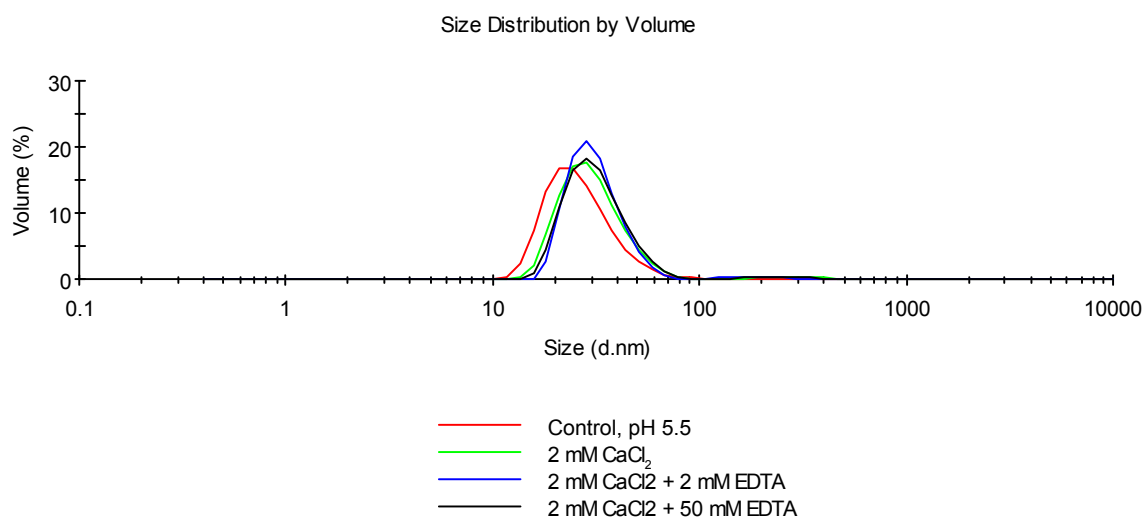


Figure 4.31: Effect of calcium and EDTA on the size distribution by volume of Mono Q-purified C1 at pH5.5. In the presence of physiological calcium concentrations, the mean size of the particles appears to increase slightly. This change is not reversible by the subsequent addition of EDTA.

Sample	PdI	Z-Average (d.nm)	Pk 1 intensity (d.nm)	Pk 2 intensity (d.nm)	Pk 1 volume (d.nm)	Pk 2 volume (d.nm)
Control	0.448	50.33	44.13	260.6	27.30	296.7
2mM CaCl ₂	0.230	126.2	43.11	252.8	31.08	268.2
2mM CaCl ₂ + 2mM EDTA	0.267	187.8	39.48	178.2	31.64	175.7
2mM CaCl ₂ + 50mM EDTA	0.238	160.5	43.43	247.5	32.34	258.7

Table 26: The effect of calcium and EDTA on Mono Q-purified C1 at pH5.5. The size distribution by volume indicates a subtle increase in the size of the C1 particles on the addition of physiological calcium concentrations at acidic pH. However, determining volume distribution relies on a number of assumptions, and may not always be reliable.

4.10.4 Melting point of Mono Q-purified C1

The final purified C1 sample was investigated by DLS in order to determine the melting point of the protein complex, and the results of this experiment are shown in Figure 4.32. The melting temperature of the C1 complex is estimated to be around 40°C, which corresponds to the melting temperature of SEC-purified C1. There appear to be three distinct size distribution profiles, according to fluctuations in the Z-average size and intensity. From 20–35°C, there are small fluctuations, while between 35–40°C these fluctuations appear to diminish. Above 45°C both size and intensity fluctuate dramatically, indicating thermal denaturation of the C1 complex. This information may be useful in designing a crystallisation strategy in the search for high resolution structures of this key immune protein.

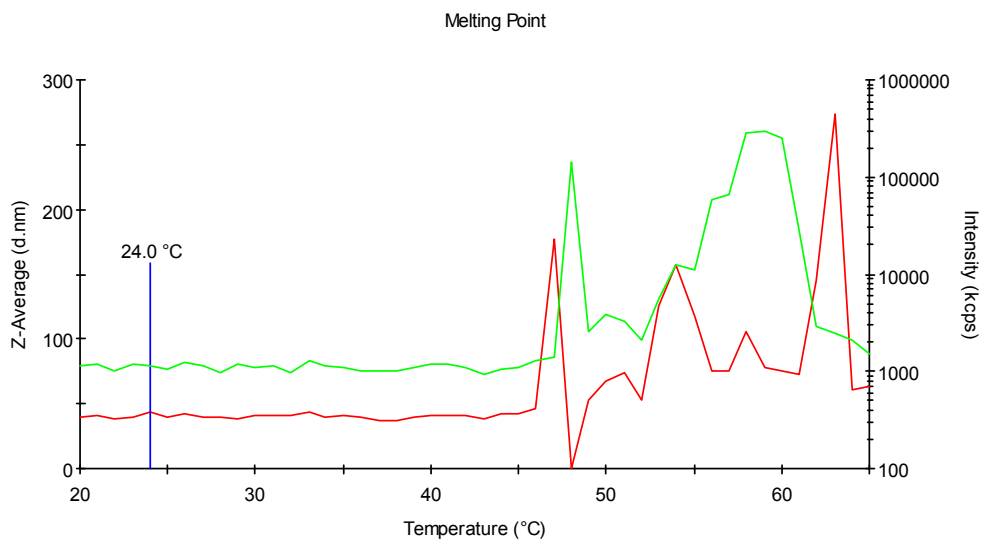


Figure 4.32: Melting point determination of Mono Q-purified C1. The Z-average (red) and intensity (green) are shown versus temperature and the melting point was determined by the ZetaSizer Nano software as 24°C. However, a melting point around 40–45°C seems like a more reasonable estimate and corresponds to the melting point determined for SEC-purified C1. There appear to be three distinct regions of the graph. From 20–35°C, there are small fluctuations in the Z-average size and intensity. Between 35–40°C, the fluctuations appear to diminish, while above 45°C both size and intensity fluctuate dramatically.

4.10.5 The effect of Mono Q-purified C1 concentration

The physiological concentration of C1 is 0.14mg/ml, while crystallisation appears to require concentrations around 1mg/ml. Mono Q-purified C1 was analysed by DLS at two different protein concentrations in order to determine the effect of varying the concentration on the size distribution of particles in the sample. This may be useful information during crystallisation trials and any further solution experiments on the purified complex. The results of these analyses are shown in Table 27.

Concentration (mg/ml)	Z-Ave (d.nm)	PdI	Pk 1 Mean Int (d.nm)	Pk 2 Mean Int (d.nm)
0.1	40.99	0.375	36.28	327.3
1	36.17	0.280	35.99	455.2

Table 27: The size distribution of Mono Q-purified C1 at different protein concentrations. At physiological concentrations in a physiological buffer the sample is polydisperse, with a Z-average size of 41nm. At 1mg/ml, the Z-average size decreases to 36nm and the PdI also decreases. Higher protein concentrations apparently cause the sample to become more homogeneous, which could be beneficial for crystallising the C1 complex.

At a protein concentration of 1mg/ml, the sample appears to be more homogeneous. This is indicated by a lower PdI and Z-average size, which closely corresponds to the results of multimodal analysis: 36.17 nm by cumulants analysis compared with 35.99 nm for the first peak in the intensity distribution. These results suggest that the high protein concentrations used in crystallisation trials may be favourable for the structural stability and homogeneity of the C1 complex.

4.11 Crystallisation of Mono Q-purified C1

Complement component C1 that had been purified by the three-step method outlined in section 4.8 was subjected to crystallisation trials. Since Structure Screen 1 & 2 had given promising results for SEC-purified C1, these were the first conditions tested. The results of these investigations are presented in the following section.

4.11.1 Pre-crystallisation test

Again, the PCT™ (Hampton Research, USA) was used to determine the best protein concentration for crystallisation trials of the C1 complex further purified by Mono Q. The results showed that the removal of contaminating proteins by IEX improves the solubility of the sample. Concentrations of 1–5 mg/ml were shown to be an appropriate starting point for these crystallisation trials, much higher than the concentrations indicated for C1 purified by SEC alone (0.25–1 mg/ml).

4.11.2 Structure Screen 1 & 2

Since Structure Screen 1 & 2 had resulted in a few promising conditions with SEC-purified C1, these were taken as a starting point for the crystallisation of the refined C1 preparation. Screens were set up using both 1 and 5 mg/ml protein concentrations. The results after one month at 4°C are summarised in Table 28. Out of a total of 96 conditions, around 40% resulted in clear drops for both protein concentrations, which is ideal. While a corresponding 60% of drops resulted in precipitate or phase separation, there was a considerable difference in the quality of the precipitates obtained with the different protein concentrations. At 1 mg/ml most of the precipitate was classified as

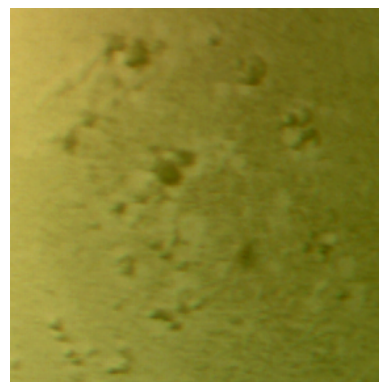
light or heavy amorphous precipitate, indicating denatured protein. In contrast, at 5 mg/ml the majority of precipitates were scored as gelatinous precipitates, which are likely to be more ordered aggregates.

	Protein concentration	
	1 mg/ml	5 mg/ml
Clear	38	37
Precipitate	45	26
Gelatinous precipitate	13	32
Phase separation	1	1

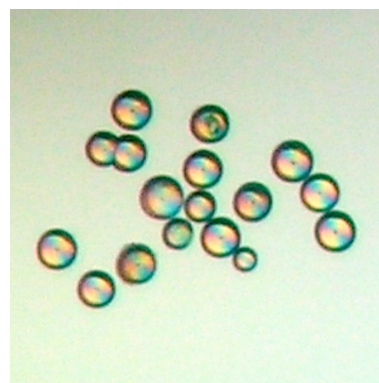
Table 28: Structure Screen 1 & 2 crystallisation data for Mono Q-purified C1. The table shows the number of drops observed in the various states described. The screens were carried out at 1 mg/ml and 5 mg/ml concentrations of the C1 complex.

In addition, three conditions were observed that appeared to show aggregates or crystals forming from precipitate at 5 mg/ml protein concentration. These are illustrated in Figure 4.33. All conditions feature a salt as the primary precipitant, which reflects the results of crystallisation trials using SEC-purified C1 (see section 4.6). The crystals were too small to be tested for diffraction, and further crystallisation of C1 was not pursued due to time constraints on the project. While it is not certain that the crystals obtained are indeed protein crystals, the results presented here will hopefully form the foundations for further experiments in the search for high-resolution structures of the native complement component C1.

0.4 M sodium tartrate



0.2M Magnesium formate



1.6 M sodium citrate, pH6.5

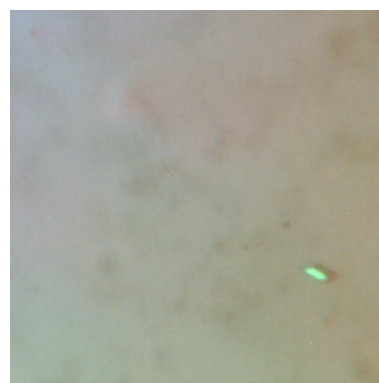


Figure 4.33: Aggregates and crystals in screens of Mono Q-purified C1. The aggregates (top), spherulites (middle) and crystal (bottom) appear to be growing from precipitated protein within the crystallisation drop. All conditions feature a salt as the primary precipitant and are at close to neutral pH.

4.12 Discussion

The aim of the work on complement component C1 was to develop an optimised novel purification scheme to isolate the native protein complex from human plasma, with an emphasis on obtaining a highly pure and homogeneous purified product for use in crystallisation trials. In addition, physiologically relevant conditions (including pH, ionic strength, and divalent metal cations) would be tested for their effect on the structure of the complex by dynamic light scattering, and the crystallisability of complement component C1 was to be explored.

Prior to the work described in this thesis, complement component C1 has been purified by exploiting its affinity for immunoglobulins or other ligands (Medicus and Chapuis, 1980; Schifferli and Steiger, 1985). For further biochemical and structural studies, individual C1 subcomponents are typically separated by adding EDTA to dissociate the complex, and C1 is reconstituted from its purified constituents (Strang et al., 1982; Tseng et al., 1997). However, the entire assembly can be precipitated from human serum using polyethylene glycol to give a functionally pure extract (Neoh et al., 1984), and this formed the basis for the novel purification scheme described in this chapter.

Complement component C1 has been successfully purified from human serum by an optimised method that does not rely on IgG binding. First, high-molecular-weight compounds are precipitated from serum using polyethylene glycol. The precipitate is then reconstituted and separated by size exclusion chromatography using Superose 6 in a physiological buffer. Finally, the peak containing C1 is collected and further purified by anion exchange chromatography using a Mono Q column. When purified by this method, C1 was shown to be pure by SDS-PAGE and dynamic light scattering.

The various peaks that eluted during the final anion exchange step of the purification were analysed by peptide mass fingerprinting and tandem mass spectrometry. These investigations served a twofold purpose: firstly, peptide mass fingerprinting allowed the various components of the C1 complex to be identified in the

corresponding samples, confirming that the purified protein was indeed C1; secondly, the analysis gave an indication of the major proteins present within each sample, and their relative abundance. Contamination with keratin from hair, skin and dust is often a problem when protein samples are analysed by mass spectrometry, and samples prepared from human serum may still contain traces of highly abundant proteins. Since the samples contain highly purified protein, peptide mass fingerprinting could be performed directly on the solution samples, thereby minimising the exposure to sources of contamination.

Various contaminants such as keratins, immunoglobulins and other abundant serum proteins were identified in both samples. However, around one third of the proteins identified were complement-related proteins. C1r and C1s were conclusively identified in both major peaks that eluted during the final purification, while C1q was only detected in the first peak. When the normalised spectrum count of C1 subcomponents in each peak is compared, we see C1 appears to be enriched in the first peak compared with the second. All other low-abundance complement-related serum proteins were more abundant in the second peak.

In combination with the SDS-PAGE analyses, this indicates that the first peak contains the intact native C1 complex in a highly purified form. However, a number of additional proteins were identified by mass spectrometry. C1-inhibitor (C1-INH) was identified in both samples, with a similar normalised spectrum count profile to that of C1r and C1s. Two further proteins could be identified with a minimum protein probability of 50% in the first sample, but not the second. These were complement factor H and C1q. Conversely, three proteins were detected in the second sample but not the first: galectin 3-binding protein, complement component C9 and mannan-binding lectin serine protease 1 (MASP-1).

The physiological role of C1-INH is to control the auto-activation of C1, and C1-INH also inhibits MASP-1 and MASP-2 (Beinrohr et al., 2008). A concentration of 22% of the normal serum value is sufficient to control spontaneous C1 activation (Windfuhr et al., 2005). The presence of C1 inhibitor in both samples suggests that at

least a portion of the purified complexes may be inhibited by binding to C1-INH. Factor H is a control protein that regulates the alternative pathway of the complement system and has been shown to interact with complement component C3. Galectin-3-binding protein has been implicated in inflammation in rheumatoid arthritis (Ohshima et al., 2003); complement component C9 is one of the terminal members of the complement cascade; and MASP-1 associates with the mannan-binding lectin of the lectin pathway of complement (see section 1.4.1, p. 9). Factor H is relatively abundant, while the remaining proteins are associated with the MBL complex, which is structurally homologous to complement component C1. Thus, while the first peak separated by the final purification step appears to contain C1, the second peak appears to contain MBL, or its sub-components.

The purified samples were further analysed by DLS, in order to determine the suitability of the purified C1 for crystallisation. Mono Q-purified C1 was shown to contain a single size distribution of particles by intensity of around 40 nm. This closely matches the proposed diameter of assembled C1 (Perkins, 1985). However, the width of the distribution suggests that some structural diversity exists within the population of purified C1 molecules. This may be due to different conformational states of the complex or the presence of contaminants of a very similar size.

Complement component C1 purified by this method appears to be relatively robust. It can be stored at 4°C for a period of one month without significant degradation in the quality of the sample. Furthermore, it does not dissociate under conditions of high ionic strength or in the presence of EDTA. This is in contrast to published reports that employ affinity purification methods.

Published reports suggest that the function of the C1 complex is affected by varying the ionic strength of the surrounding medium (Lepow et al., 1958; Colten et al., 1968). The authors propose that the observed loss of activity is caused by dissociation of the complex at ionic strengths outside the range of 0.1–0.15. However, the results outlined on page 215 suggest that high and low ionic strengths do not dissociate C1. Instead, a small increase in the hydrodynamic radius of the complex is observed at salt

concentrations outside the physiological range at pH 7.4. This increase in the size of the C1 complex likely results from conformational changes within the molecule, which could explain the loss of function at high and low ionic strength described by Lepow et al., 1958 and Colten et al., 1968. Furthermore, differences in the structural profile of the purified C1 complex were observed at high and low ionic strength: at low ionic strength, C1 appears to have a tendency towards aggregation, while at high ionic strength only the shape of the molecule appears to be affected. This suggests that different mechanisms may be responsible for the loss of function at low ionic strength compared to the loss of function at high ionic strength.

Overall, the findings presented here support the idea that C1 is only functionally active within a narrow range of ionic strengths, corresponding to the physiological range of human blood plasma. While the biological implications of these findings are unclear, there may be more pertinent implications for the crystallisation of the C1 complex. For example, moderate salt concentrations are required for the complex to adopt a native state, while elevated concentrations such as those used in crystallisation may affect the conformation adopted by the molecule and alter it so that it does not accurately reflect the physiological molecule.

EDTA is commonly used to dissociate the individual C1 subcomponents after purification by affinity methods. Here the C1 complex was purified by the novel method combining PEG precipitation with gentle size exclusion chromatography, and the effect of EDTA on the resulting C1 was evaluated. Adding EDTA to Mono Q-purified C1 causes a small amount of larger aggregates to form. At EDTA concentrations higher than those typically required to dissociate the complex (see for example Lepow et al., 1963), 0.75 nm particles are observed in addition to the 40 nm C1 particles. These correspond to a predicted molecular weight of 1.7 kDa, which is extremely small, and therefore does not suggest that C1 is dissociating into its individual subcomponents. These observations are in stark contrast to published literature on the subject, which shows that the complex is disassembled in the presence of EDTA, since this sequesters the calcium required by the C1 for structural stability.

One possible explanation for this inconsistency can be found in the different methods of C1 preparation. As explained in section 1.7.1, p. 37, previous purification protocols have relied on the affinity of C1 for IgG in order to selectively separate the complex from contaminating proteins. Typically, the individual sub-components are further purified by dissociating the complex via the addition of EDTA, and the C1 is finally reassembled from its pure constituents. However, in the method presented here the complex is gently precipitated from human serum — along with other high-molecular-weight proteins — and further purified in physiological or near-physiological buffers. The process of binding and release from IgG during affinity purification methods makes C1 susceptible to dissociation by EDTA whereas C1 dissociation is apparently prevented when C1 is purified by gentle methods. This suggests that some factor may be associated with C1 that prevents dissociation by EDTA and is removed during binding of the C1 complex to IgG and other ligands.

Peptide mass fingerprinting analysis of purified C1 showed that C1-INH was also present in the sample. During complex formation, C1 inhibitor dissociates C1r and C1s from the activated C1 macromolecule. However, it may also suppress spontaneous activation of C1 via a reversible interaction with C1r and C1s (Davis et al., 2008). Interestingly, C1 inhibitor also regulates lectin pathway activation via inactivation of MASP-2 (Kerr et al., 2008). This could explain why C1-INH is found in both purified C1 and also in the sample thought to contain MBL, the homologue of C1 in the lectin pathway.

If IgG is used to purify the complex, either the inhibitor is likely to be displaced, or the method will simply not purify inhibitor-bound C1 complex. The presence of C1-INH in the sample of C1 purified by the novel method described here suggests that it may be bound to C1. Furthermore, this interaction may be causing the complex to resist disassociation by EDTA. It is possible that binding C1-INH alters the affinity of the C1 complex for calcium or otherwise protects the bound metal ions required for structural stability from being sequestered by EDTA.

The results presented on page 188 suggest that purified C1 may form large aggregates in the presence of physiological copper concentrations at low pH. Vigorous aggregation of serum amyloid P component also occurs in the presence of copper at acidic pH (Potempa et al., 1985). Since the copper concentrations used in those studies were much higher than physiological copper concentrations, the authors considered the solubility characteristics of copper at acidic pH values. However, no insoluble copper complexes in the absence of SAP were identified in control studies. The authors concluded that local environments enriched in copper at acidic pH — such as sites of infection and inflammation — may trigger SAP auto-aggregation and potentiate SAP binding to those ligands (e.g. C4-binding protein) for which SAP aggregation is a requirement.

It is possible that a similar mechanism of aggregation exists for C1, which may have some relation to its function *in vivo*. Studies carried out in our laboratory by Katy Ginger suggest that both copper and zinc specifically interact with C1 and cause inhibition of the protease activity of the complex. However, whether these are exclusively due to C1, or to what extent MBL and/or other co-purifying proteins contribute to these observations is unclear. However, the interaction of both complement component C1 and SAP with copper and calcium under physiological and pathological conditions is clearly a phenomenon that merits further exploration.

Initial investigations into the crystallisability of the complex have been carried out in the search for high-resolution structures by X-ray crystallography. Protein concentrations of 1–5 mg/ml were shown to be an appropriate starting point for the crystallisation of complement component C1. The concentration of protein required for crystallisation is inversely proportional to its molecular weight and so a large protein such as C1 may be expected to crystallise at these lower concentrations.

Purified C1 was subjected to crystallisation trials using Structure Screen 1 & 2, since these had shown promising results using C1 samples at intermediate stages of purification. Around 60% of the conditions tested resulted in precipitate or phase separation, and there was a considerable difference in the quality of precipitates

obtained with the different protein concentrations. At 1 mg/ml most of the precipitate showed denatured protein, while at 5 mg/ml the majority of precipitates showed more ordered states, including aggregates apparently growing from precipitate, spherulites, and small crystals. These promising conditions all feature a salt as the primary precipitant, suggesting that electrostatic interactions may be stabilising the C1 complex. By contrast, precipitates were obtained in many conditions using PEG as the primary precipitant.

These observations are not surprising, considering that low concentrations of PEG are used to precipitate the complex from serum. Furthermore, the experiments described in section 4.10.1, p. 215 show that high ionic strength induces a small increase in the hydrodynamic size of the C1 complex, instead of causing dissociation as previously assumed. In addition, the results presented in section 4.10.5, p. 225 show that the C1 population becomes more homogeneous at higher protein concentrations, such as those used during crystallisation. This may be a consequence of the equilibrium between any free subunits and the C1 complex: in accordance with Le Chatelier's principle, increasing the protein concentration will shift the position of equilibrium towards the state of lowest molecularity, in this case the assembled complex. Thus the combination of high protein concentration and ionic precipitants appear to offer an environment conducive to the crystallisation of complement component C1.

Another variable to consider in the future search for C1 crystals would be temperature. As described in section 4.10.4, p. 224, C1 is affected by temperature. Below 35°C there are small fluctuations in the Z-average size and intensity by DLS, while between 35–40°C these fluctuations appear to diminish. Above 45°C both size and intensity fluctuate dramatically, indicating thermal denaturation of the C1 complex. These results suggest that C1 is most stable within the narrow range of temperatures corresponding to its physiological milieu, at around 37°C. The crystallisation attempts described here were carried out at 4°C, since the low temperature allows a gradual equilibration of the crystallisation drop with the reservoir

solution and may slow down any undesirable auto-activation of the complex. However, given the decreased stability of the complex at temperatures below around 35°C, better results may be achieved at higher crystallisation temperatures.

The crystals obtained during the initial crystallisation trials were too small to be tested for diffraction, so there is considerable scope for further work on the crystallisation of complement component C1. In addition to expanding the scope of crystallisation trials as discussed above, the purified C1 sample may yet contain other complement-related proteins and high-abundance serum proteins, so further purification or alternative purification methods could also be explored. Mono Q-purified C1 may also be a more suitable sample than SEC-purified C1 for electron microscopy. Finally, the interaction of purified C1 with divalent metal cations — and also C1 inhibitor — merit further investigation. However, the aims set out for this project have been adequately fulfilled.

A novel purification scheme to isolate native complement component C1 from human plasma was developed, using a combination of PEG precipitation, size exclusion chromatography and anion exchange chromatography. All of the C1 subcomponents were shown to be enriched in the purified sample by SDS-PAGE and mass spectrometry. Dynamic light scattering studies showed a single population of particles around 40nm in diameter, which corresponds to the estimated size of the C1 complex. Therefore C1 has been purified by a new procedure in a form suitable for crystallisation. In addition, investigations into the effects of relevant metal ions on the structure of the purified complex under physiologically relevant conditions were pursued. Finally, the crystallisability of complement component C1 was explored in the hope of achieving usable crystals. Several promising conditions were found, and this project therefore reached a natural conclusion in accordance with its objectives and the available time and resources.

Chapter 5

Investigating the ligand binding properties of SAP

5.1 Introduction

Serum amyloid P component is a member of the pentraxin family, which also includes the acute phase protein, C-reactive protein (CRP). Pentraxins are highly conserved and are characterised by their calcium-dependent ability to bind a variety of ligands. Interestingly, SAP activates complement by binding to C1q (Ying et al., 1993).

In human serum, SAP exists as a pentamer (Hutchinson et al., 2000), while in the presence of EDTA, SAP forms stable decamers composed of two pentamer rings interacting face to face (Pepys et al., 1997). X-ray and neutron scattering data showed that SAP pentamers were reproducibly obtained in the presence of calcium at pH 5.5 or in the presence of MO β DG and calcium at pH 6.0–8.0, while SAP decamers were obtained in the presence of EDTA between pH 5.5–8.0 (Ashton et al., 1997). In the presence of excess calcium, isolated human SAP auto-aggregates and precipitates from solution. This calcium-dependent self-assembly of SAP pentamers into larger aggregates is inhibited by all known ligands (Thompson et al., 2002).

The secondary structure of human SAP and CRP in the presence of calcium, magnesium, and phosphorylcholine has also been investigated using Fourier transform infrared spectroscopy (Dong et al., 1994). Significant differences were noted in the response of SAP compared with CRP. Calcium-dependent changes were observed in the secondary structure of SAP, but the spectrum was not affected by magnesium. Furthermore, phosphorylcholine in the presence of calcium did not affect the FTIR spectrum of SAP. These findings suggest that members of the pentraxin family — such as SAP and CRP — display calcium-dependent secondary structure conformational changes but differ in their responses to phosphorylcholine and magnesium.

These properties are thought to be relevant to the incompletely understood biological function of these highly conserved proteins, which includes activation of the complement system and association with amyloid deposits. The aim of the work on

SAP was thus to perform preliminary studies on the effect of metal cations on the ligand-binding and structural properties of SAP under physiologically relevant conditions.

5.2 Purification of SAP

The purification scheme used to prepare SAP was adapted from that of De Beer and Pepys, 1982, using ion exchange chromatography instead of size exclusion as the final chromatography step. This method of preparing SAP relies on its calcium-dependent affinity for agarose and is described fully in section 3.2, p. 131. Briefly, human plasma is incubated with an agarose-based gel filtration medium in the absence of any additional calcium (the calcium naturally bound to SAP being sufficient for the interaction with agarose to occur). Any SAP molecules in the sample bind to the medium, while unbound contaminants can be washed away. Finally, SAP is eluted with excess EDTA, which sequesters calcium ions and thus disrupts the binding of SAP to the column. The eluted fractions still contain some contaminants, but these can be removed by IEX using a MonoQ column and NaCl gradient to leave a single 25kDa protein — pure SAP — on SDS-PAGE gels as shown in Figure 5.1.

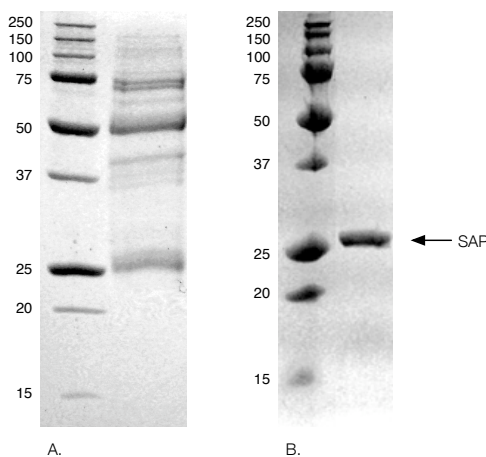


Figure 5.1: SAP in various stages of purification. A. Fraction eluted from a Sepharose 4B column. Numerous bands can be seen, indicating that the sample is still relatively impure. B. Fraction eluted after further purification using a MonoQ column. A single 25kDa band is clearly visible, indicating that SAP is present in a highly purified form.

Purified SAP was exchanged into an EDTA-free buffer (10mM Tris, pH7.4, 150mM NaCl) and analysed by DLS to determine the size distribution of particles in the sample. The results are summarised in Figure 5.2 and Table 29. The sample was shown to contain a single species of approximately 10nm in diameter according to the intensity distribution. In the presence of EDTA, this corresponds to a molecular weight of around 250kDa, although in the absence of EDTA, the molecular weight is apparently closer to 150kDa. This shows that in the absence of EDTA or metal ions, purified SAP is present as a 125kDa pentamer in solution. The difference in the observed size by intensity can be explained through the shape effects of the ring-like SAP structure on the apparent hydrodynamic radius. SAP decamers, which consist of two pentamer rings stacked face to face, have a more globular shape that more closely corresponds to the spherical approximation used in molecular weight calculations.

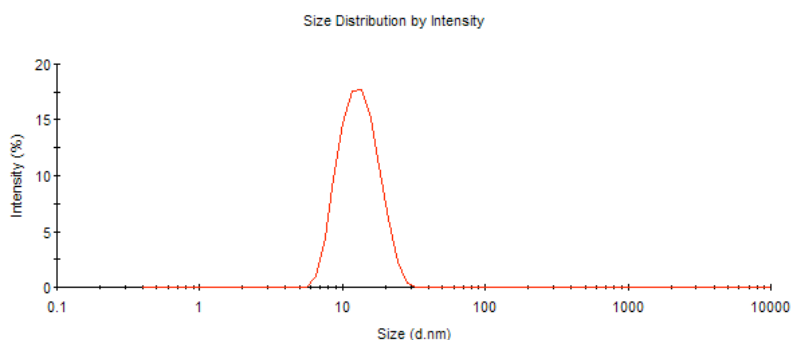


Figure 5.2: Size distribution by intensity of purified SAP by DLS. The figure shows a monomodal distribution of particles approximately 10 nm in diameter. The concentration of SAP was 10 μ M (10 mM Tris, pH 7.4, 150 mM NaCl, 50 mM EDTA).

EDTA	Size (d.nm)	% Pd	MW (kDa)
+	12.62	40.9	251
-	10.24	32.5	154

Table 29: Measurements derived from the DLS analysis of purified SAP. The estimated size, percentage polydispersity and molecular weight are indicated. The results show that SAP is a decamer in the presence of EDTA, and a pentamer in 10 mM Tris, pH 7.4, 150 mM NaCl.

5.3 Aggregation assays

In the absence of calcium, SAP forms stable decamers composed of two pentamer rings interacting face to face (Pepys et al., 1997). In the presence of excess calcium, isolated human SAP auto-aggregates and precipitates from solution. This calcium-dependent self-assembly of SAP pentamers into larger aggregates is inhibited by all known ligands (Thompson et al., 2002). One ligand with particular physiological relevance is phosphatidylethanolamine (PE).

This phospholipid is usually found in the inner leaflet of the plasma membrane, along with phosphatidylserine (PS) and phosphatidylinositol. In contrast, the outer leaflet typically features phosphatidylcholine, sphingomyelin and a variety of glycolipids. However, during the early stages of apoptosis, this asymmetry breaks down as PE and PS flip-flop from one side of the membrane to the other (Familian et al., 2001). Thus PE may be implicated in the biological function of SAP. In order to investigate the interactions between calcium, SAP and PE, aggregation assays were performed and monitored by DLS.

Under physiological conditions, SAP forms pentamers with a molecular weight of 125kDa. When calcium is added, two distinct peaks are observed in the particle size distribution determined by DLS (see Figure 5.3). The smaller size peak is representative of pentameric SAP, while the larger size peak is caused by aggregation of SAP pentamers via AB face-to-face stacking mediated by bound calcium and Glu-167 (Thompson et al., 2002). The calcium-dependent aggregation of SAP over time is shown in Figure 5.3. When SAP is calcified and then PE is added to the sample, the aggregation of SAP continues, but at a reduced rate (Figure 5.4). However, when PE is added to SAP first and the sample is subsequently calcified, no aggregation is observed (Figure 5.5). These results clearly demonstrate the inhibition of calcium-dependent aggregation of SAP by its binding of the ligand PE.

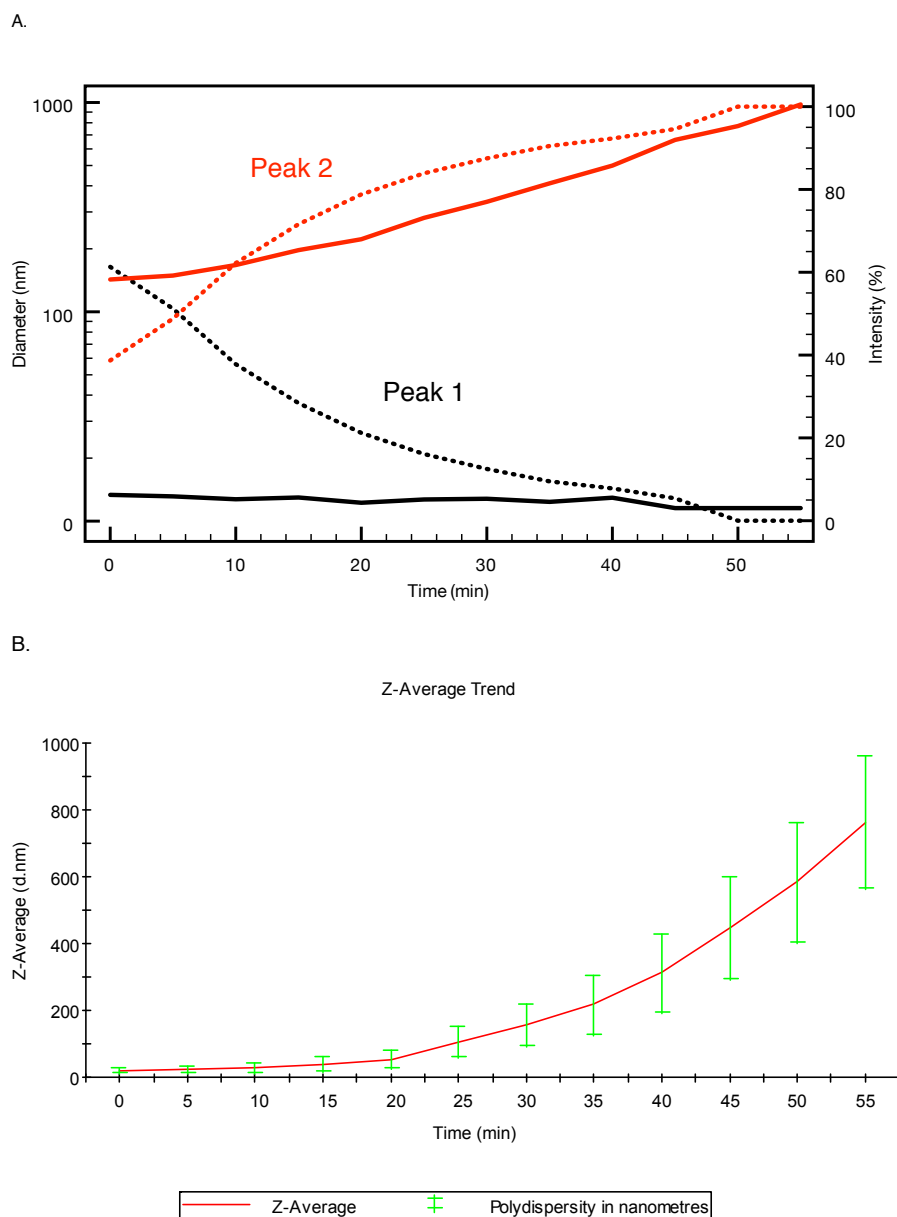


Figure 5.3: Aggregation of SAP monomers on the addition of calcium. This figure shows time-courses of DLS measurements from the two major peaks in the size distribution profile of SAP after calcium is added to the sample. Physiological concentrations of SAP (50 $\mu\text{g/ml}$) were prepared in 10mM Tris pH8, 140mM NaCl and the sample was calcified with 1mM calcium, pH7.4. DLS measurements were taken at 5 min intervals from the introduction of the calcified sample into the machine. A. Two distinct peaks are observed in the size distribution: peak 1 corresponds to SAP, while the second peak is indicative of SAP aggregates. For each peak the particle size (solid line) and intensity (dashed line) are shown. The figure shows the disappearance of SAP pentamers (peak 1) and the appearance of aggregates (peak 2) after the addition of excess calcium. B. Aggregating SAP shows increasing polydispersity with time. This likely reflects the increasingly heterogeneous population of aggregates.

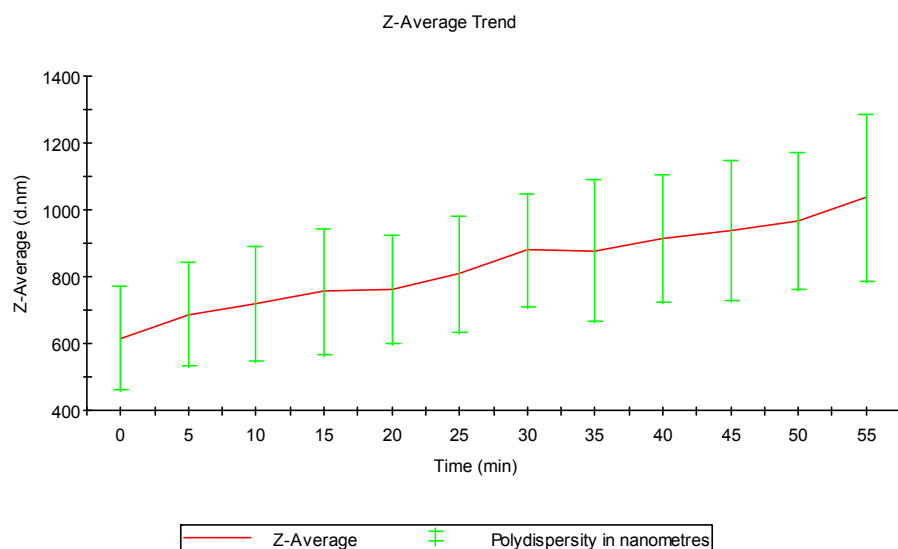


Figure 5.4: Adding PE to calcified SAP does not inhibit the further aggregation. Physiological concentrations of SAP (50 $\mu\text{g}/\text{ml}$) were prepared in 10mM Tris pH8, 140mM NaCl. After adding 1mM calcium, the sample was incubated at room temperature for one hour to allow aggregates to form. Finally, PE was added to a final concentration of 1mM and DLS measurements were taken every 5 minutes. The plot shows that adding PE to the calcified sample does not inhibit the calcium-dependent aggregation of SAP.

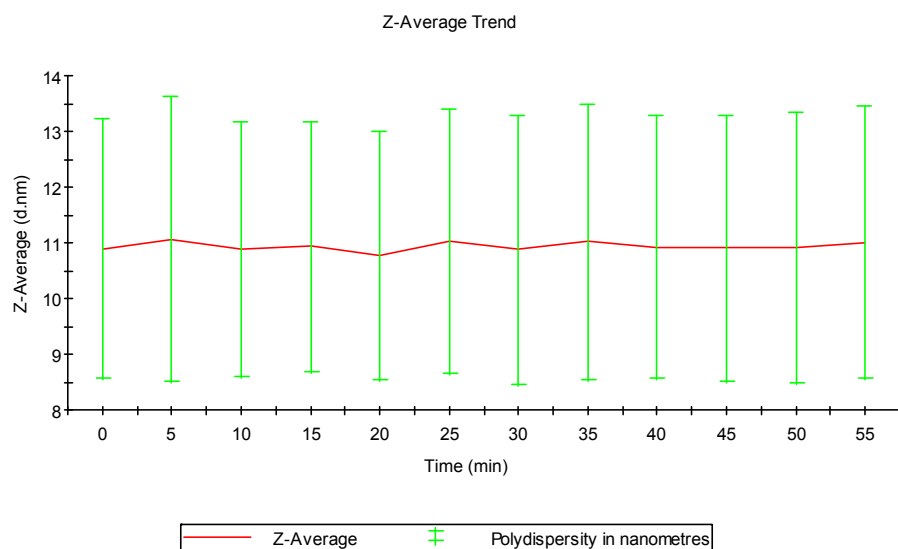


Figure 5.5: PE prevents calcium-induced SAP aggregation. Physiological concentrations of SAP (50 $\mu\text{g}/\text{ml}$) were prepared in 10mM Tris pH8, 140mM NaCl. In this experiment, PE was added first to a final concentration of 1mM, followed by the same concentration of calcium before DLS measurements were taken every 5 minutes. The results show that when PE is present before calcium is introduced, the calcium-dependent aggregation of SAP is inhibited by PE.

5.4 Discussion

The experiments described in section 5.3 provide a simple and rapid method for determining the effect of metal- and ligand-binding by SAP. These formed the basis for further studies carried out by other members of our laboratory, and this work is currently ongoing. The findings presented in this thesis, and their implications within the context of the investigations carried out by others, will be discussed in this section.

Serum amyloid P component has been found to associate *in vitro* with PE in the experiments described in section 5.3, and with a variety of other polysaccharides and proteins in the presence of calcium at neutral pH in the published literature. Thus the method for assaying the effect of divalent metals on ligand-binding and auto-aggregation described here may be extended to other ligands. The role of divalent cations and acidic pH on the binding of SAP to the ligand zymosan has been previously investigated by Potempa et al., 1985, since they may play a role in the biological function of the molecule. The authors reported that binding occurs not only in the presence of calcium but also in the presence of copper and zinc, and that SAP did not bind the ligand in the absence of added metal or in the presence of various other metal cations that may be present in normal human blood plasma.

Whereas calcium-mediated SAP-zymosan binding decreases at low pH, copper-mediated binding increases substantially. Thus, in addition to calcium at neutral pH, both copper and zinc — at neutral and in particular at acidic pH values — mediate SAP binding to polysaccharide ligands. This suggests that SAP may well be considered a copper- as well as a calcium-dependent protein, and that this reactivity is favoured under those conditions of lowered pH which may result from metabolic processes occurring at local sites of inflammation (Potempa et al., 1985). Similarly, copper-dependant conformational changes and a modulation of the protease activity of the complex by different metal cations were observed in SAP in work carried out by Sasha Derrington in our laboratory.

Serum copper levels increase during inflammation, most notably in association with elevation of ceruloplasmin, a copper-transport protein that donates copper to enzymes such as cytochrome oxidase. This process is greatly enhanced in an acidic environment (Owen, 1975), which is a feature of sites of infection and inflammation. Furthermore, a selective precipitation between SAP and ceruloplasmin was identified by Uhlenbruck et al., 1978. It was proposed that at low pH, a copper exchange reaction could thus occur between ceruloplasmin and SAP (Potempa et al., 1985).

Experiments performed in our laboratory by Sasha Derrington showed that SAP and ceruloplasmin formed stable complexes, suggesting that a copper exchange reaction may take place between the two proteins. Clearly, the role of ceruloplasmin and copper under physiological and pathological conditions in the structure and function of SAP is a matter that remains to be fully explored. The results presented here may offer a rapid and simple method for assaying the metal- and ligand-binding properties of serum amyloid P component, and further studies examining the role of these metals and pH could contribute to understanding the physiological role of SAP.

At an SAP concentration of 1 mg/ml — in the range of concentrations required for crystallisation (Wood et al., 1988) — 1 mM calcium was sufficient to cause rapid SAP auto-aggregation. However, by keeping the concentrations of both protein and metal sufficiently low, auto-aggregation of SAP can be avoided (Potempa et al., 1985). The results presented in section 5.3 show that calcium-dependent aggregation could be detected by dynamic light scattering even at physiological SAP concentrations in the presence of 1 mM calcium. These observations may be consistent with the findings described by Potempa et al., 1985, since dynamic light scattering allows the detection of even small numbers of aggregated particles, whereas these may not have been observed in the original study.

Vigorous aggregation of SAP also occurs in the presence of copper concentrations above 0.5 mM at acidic pH (Potempa et al., 1985). The authors concluded that local environments enriched in copper at acidic pH — such as sites of infection and inflammation — may trigger SAP auto-aggregation and potentiate SAP

binding to those ligands (e.g. C4-binding protein) for which SAP aggregation is a requirement. The results of these auto-aggregation experiments have implications for the crystallisation of SAP, which readily forms crystals in the presence of calcium and PEG 6000 at pH 5.5. Interestingly, a copper-dependent aggregation at acidic pH is also observed with complement component C1 (p. 188 for results and p. 234 for discussion).

When all the available information is taken into account, the picture that emerges is that copper and zinc may replace the calcium in the binding pocket of SAP (and possibly also in calcium-binding sites on C1). The K_d of calcium binding to SAP has been estimated at 0.1 μ M and 1.5 mM for each of the two binding sites respectively (Heegaard et al., 2006). Thus, calcium II is held much more loosely by the molecule than calcium I. When all of these observations are taken together, they suggest that at least one of the calcium-binding sites on SAP may instead be occupied by a copper, zinc or other divalent metal ion under certain conditions.

This opens up the possibility of crystallising SAP bound to different divalent metal ions, and co-crystallising SAP with different divalent metal ions and ligands. If SAP can be successfully crystallised with copper in place of calcium then near-UV CD spectroscopy or X-ray absorption spectroscopy (EXAFS) could be used to study the local geometry of the co-ordinated metal ion. Particle-induced X-ray emission (PIXE/microPIXE) can provide a quantitative analysis of the metal content of a protein sample, which could give information about the occupancy of the metal site. Finally, fibre diffraction and TEM could be explored as a means of studying the stacks of aggregated SAP pentamers that form in the presence of calcium and copper.

Chapter 6

Validation of the ZiCo design

6.1 Introduction

As described in section 1.6, ZiCo is a small peptide designed to switch between an oligomeric coiled coil and a zinc finger fold on binding zinc. The design was used as a template to create a family of peptides based on the same principles, incorporating various sequence features in order to determine their effects on folding and stability. Two members of this family were obtained from Prof. Dek Woolfson: the original design, ZiCo1; and the modified peptide ZiCo95.

The differences between the two peptides are illustrated in Figure 6.1. ZiCo1 contains four histidine residues while the ZiCo95 sequence displays the classical zinc finger motif, containing two cysteine and two histidine residues. This renders ZiCo95 more vulnerable to oxidation and the formation of disulphide bonds within or between peptide molecules. Other modifications include the addition of glutamate residues, making ZiCo95 much more negatively charged at a given pH than ZiCo1 (as illustrated by the peptide charge calculations in Appendix II, section i).

A.	1	10	20
B.	-YIHALHRKAFAK	IARLERHIRALEHAA	
C.	EYEC	CLKCEIEFLKNENLKLHQKIH	ALEQ

Figure 6.1: A comparison of the ZiCo1 and ZiCo95 sequences. A. Residue number. B. ZiCo1 sequence. C. ZiCo95 sequence. Residues involved in zinc binding are shown in yellow and identical residues are shown in black. The key modifications involve the substitution of cysteines for two of the histidine residues, and the incorporation of glutamate residues.

The overall goal of this project was to develop a crystallisation strategy for both the metal-free oligomer and the zinc-bound monomer forms of both peptides that would lead to high-resolution structures. These are valuable in terms of testing our understanding of the rules that govern protein folding, which has practical implications for designing proteins with novel functions for use in industry and medicine.

6.2 Crystallisation of ZiCo1 and ZiCo95

Prior to the work described here, diffracting crystals of ZiCo1 had been grown by David Papostolou using a peptide concentration of 2.25mM ZiCo and 0.4M LiSO₄ as the precipitant at near neutral pH in the presence of 1mM zinc chloride. However, all attempts to solve the structure using MR had proved unsuccessful. It was thus decided that the screening of new conditions to complement those already tested could be valuable in terms of yielding alternative crystal forms from which a structure solution might be found.

In designing a crystallisation strategy for the two ZiCo peptides, a number of factors were considered. Firstly, many of the standard commercially available screens are optimised for globular proteins, using conditions that have been shown to successfully yield crystals. However, ZiCo does not fit the typical profile for a globular protein because of its small size and structural flexibility. It has been shown to be highly sensitive to varying buffers, pH, zinc and salt concentrations (Eleonora Cerasoli, personal communication). In light of this fact, a systematic approach was favoured over the sparse-matrix style screens used in the previous crystallisation attempts, in the hope that this would give a clearer understanding of the solubility and crystallisability of the ZiCo peptides. The Stura Footprint screen (Molecular Dimensions, UK) was chosen as a starting point to explore the crystallisation conditions of ZiCo1 and ZiCo95 because it allows systematic screening of the protein solubility curve against commonly used precipitants at basic, acidic and neutral pH.

6.2.1 ZiCo95 crystallisation attempts and analysis

One of the first steps in the endeavour to crystallise ZiCo95 was to establish an appropriate concentration of both peptide and zinc in the crystallisation solution. Increasing the concentration of zinc ions should drive the equilibrium towards the zinc-bound monomer, which could be an important factor in any crystallisation attempts. The Pre-Crystallisation TestTM (Hampton Research, USA) was used to identify a suitable peptide concentration and peptide:zinc ratio for the initial screening as described on page 133. The results of the PCTTM indicated that 0.5mM peptide would be a suitable concentration for screening purposes. A positive result was obtained with 1:0, 1:1, 1:5 and 1:10 peptide:zinc ratios using zinc chloride, so these conditions were used to set up the initial screens.

6.2.1.1 Crystallisation trials

Numerous attempts were made to crystallise ZiCo95, using the Stura Footprint Screen (Molecular Dimensions, UK) as a starting point. While none of the initial screens gave any crystals, the systematic nature of the screens meant that promising conditions could be identified for further exploration. Such conditions were identified where an increase in precipitant concentration resulted in a transition from a clear drop to precipitate. The sudden phase transition suggests that the conditions border a region of the protein solubility curve that includes the nucleation and metastable zones required for crystallisation. Finer sampling of this region of crystallisation space may yield the right conditions for protein crystals to form and grow.

Peptide:zinc	Precipitant	Buffer	pH
1:1	15–24 % PEG 600	0.2 M imidazole malate	5.5
1:1	20–25 % PEG 4 K	0.2 M imidazole malate	7.0
1:1	17.5–22.5 % PEG 10 K	0.2 M imidazole malate	8.5
1:1	< 0.75 M ammonium sulphate	0.15 M sodium citrate	5.5
1:1	< sodium/potassium phosphate	n/a	7.0
1:1	< 1 M citrate	10 mM sodium borate	8.5

Table 30: Promising conditions from initial ZiCo95 crystallisation screens. This table shows conditions that resulted in a transition from a clear drop to precipitate in the initial screen using 0.5 mM ZiCo95 and a range of peptide:zinc ratios. These promising conditions were then expanded into further screens by varying the precipitant concentration within the identified range and the pH by ± 1 unit around the identified condition.

Table 30 shows the conditions that were used as starting points for further exploration of the crystallisability of ZiCo95. By varying the precipitant concentration and pH, the promising conditions from the initial screens were expanded into further 4×6 screens.

Since cysteines are present in the ZiCo95 sequence, the reducing agent BME was included as a parameter in the crystallisation trials. In addition, experiments were set up using different peptide concentrations (0.3, 0.5 and 0.8mM), crystallisation buffers (150mM Tris or 150mM phosphate) and temperatures (4°C and room temperature). Despite extensive attempts to crystallise ZiCo95, none of the conditions tested produced crystals of the peptide.

6.2.1.2 SEC analysis

In order to further understand the solution behaviour of ZiCo95, samples were analysed by SEC in the presence and absence of various additives. The column was calibrated according to the manufacturer's instructions and the separations were performed in 50mM Tris, pH7.5, 50mM sodium chloride, using 20nmol of peptide per run. Multiple peaks were seen in the chromatogram both with and without an equal concentration of zinc chloride in the sample and elution buffers, as shown in Figure 6.2. When the peptide was reduced by the inclusion of BME in the sample buffer, ZiCo95 eluted as a large single peak with a predicted molecular weight of 7.9kDa, together with a small amount of higher molecular weight material both with and without zinc (see Figure 6.3).

The results of the SEC analysis show that ZiCo95 exists as a heterogeneous population of oligomeric species in solution, in both the presence and absence of zinc. The zinc-free peptide appears to be relatively uniform in terms of the number and type of oligomers observed: most of the peptide molecules assemble into trimers — as designed — with only a small proportion forming tetramers. In the presence of zinc, ZiCo95 displays a much more heterogeneous population of oligomeric states. Most of the peptide is found as pentamers, followed by trimers and finally a small proportion of monomers. This behaviour may explain the lack of success in crystallising ZiCo95, since a key requirement for successful crystallisation is generally a homogeneous population of protein molecules. As Figure 6.2 shows, ZiCo95 forms multiple oligomeric species in both the presence and absence of zinc. Increasing the peptide concentration is likely to shift the balance of equilibrium towards oligomeric states, contributing to heterogeneity.

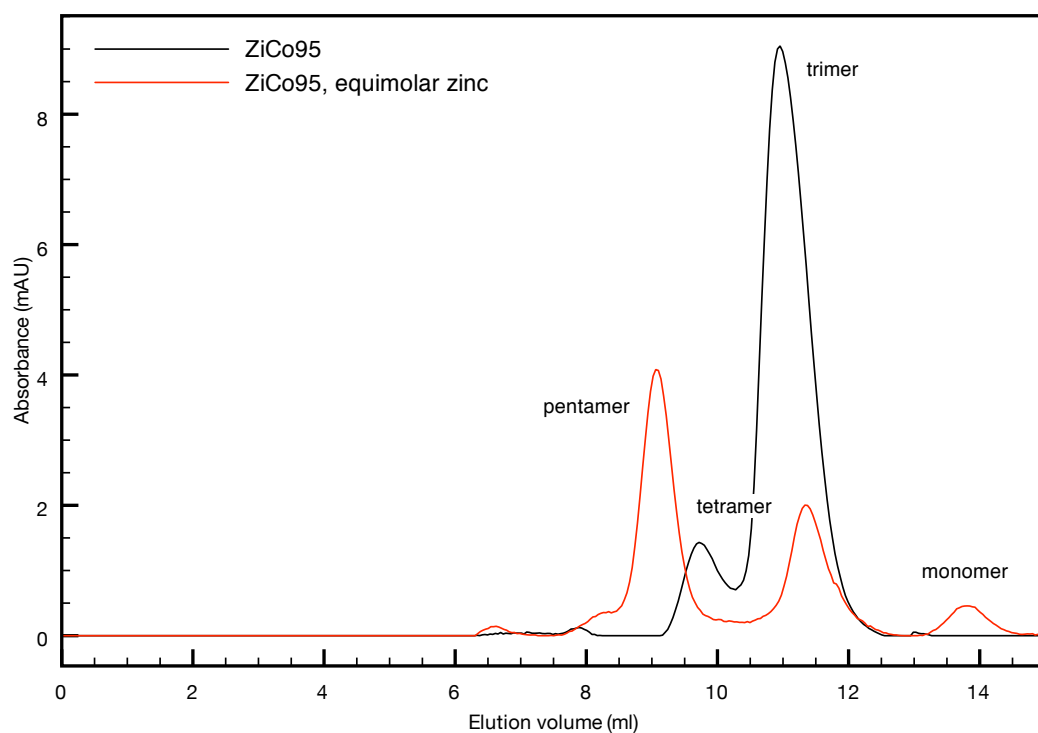


Figure 6.2: Elution profile of ZiCo95 in the presence and absence of equimolar zinc chloride. Various oligomeric states can be distinguished and identified by their calculated molecular masses. The results show that in the absence of zinc, ZiCo95 exists primarily as a trimer, with a small amount of a tetrameric species. In the presence of equimolar zinc, a small amount of monomer is observed, with large amounts of trimeric and tetrameric assemblies.

Results subsequently obtained from the Woolfson laboratory show that in the presence of equimolar zinc, a significant portion of ZiCo95 is bound to two zinc ions. At high concentrations, such as those used in the crystallisation trials, the peptide precipitates spontaneously from solution. This is likely to be due to cross-linking of the peptides via bound zinc, suggesting that it may be impossible to crystallise ZiCo95 in its zinc-bound form. In light of these findings, and due to time constraints, further attempts to crystallise ZiCo95 were not pursued.

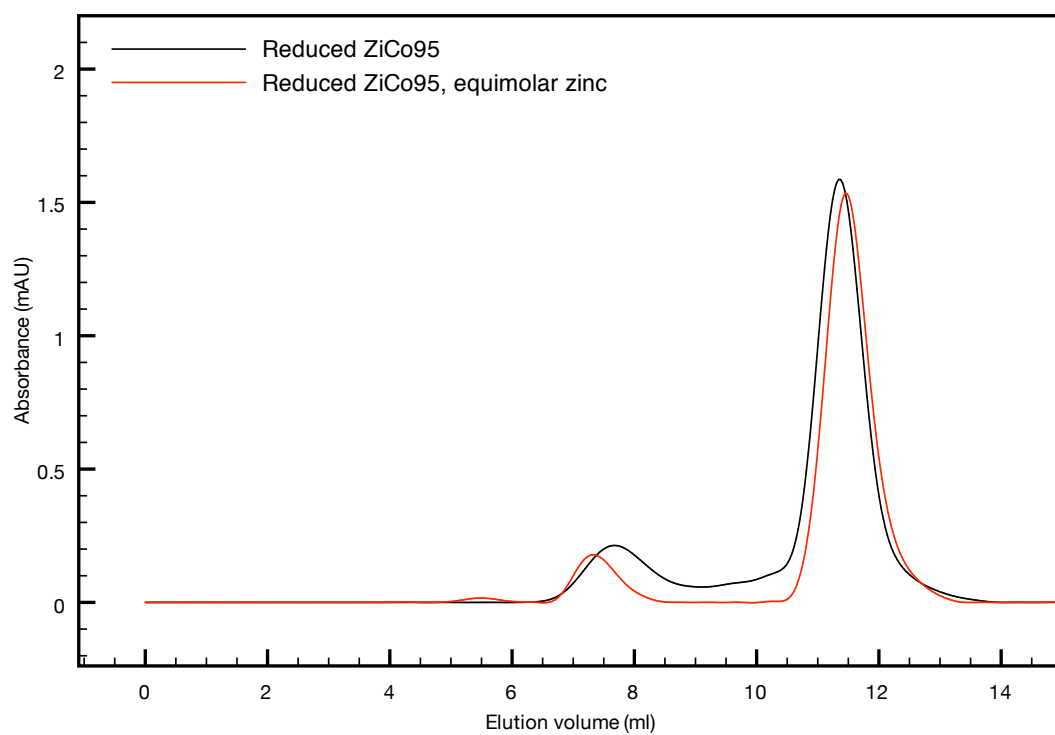


Figure 6.3: Elution profile of reduced ZiCo95 in the presence and absence of equimolar zinc chloride. The peptide elutes as a large peak with a predicted molecular weight of 7.9 kDa, together with a smaller peak corresponding to an aggregated state. The same elution profile is observed both in the presence and absence of equimolar zinc.

6.2.2 ZiCo1 crystallisation and optimisation

ZiCo1 was obtained lyophilised and frozen at -20°C from Eleonora Cerasoli (Sussex, UK) and Beth Bromley (Bristol, UK). Stock solutions of 5–10 mg/ml were prepared by dissolving the peptide in water, and stored at 4°C . The Pre-Crystallisation Test (Hampton Research, USA) indicated that a 1 mM solution of the peptide in 10 mM phosphate, pH 7.5, 50 mM NaCl and an equimolar concentration of zinc were suitable for crystallisation purposes.

6.2.2.1 *Crystallisation trials*

The initial systematic screens produced a number of crystals when zinc was included in the sample solution, and when both PEGs (Table 31) and salts (Table 32) were used as the primary precipitant. The results show the stark differences in crystal morphology obtained when using salts as opposed to PEGs. Conditions containing PEGs gave rise to numerous small, irregular crystals, while conditions with salts produced larger, single crystals. ZiCo1 crystallises at varying pH values, with the best crystals obtained between pH 7 and 8.5. This corresponds to the pH at which the peptide has been characterised in solution, so any structural information derived from these crystals is likely to be relevant to validating the design of this novel peptide.

Table 31: Initial hits from PEG conditions of the Stura Footprint screen


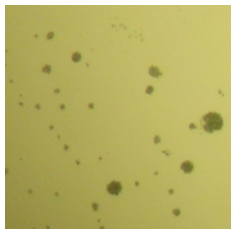
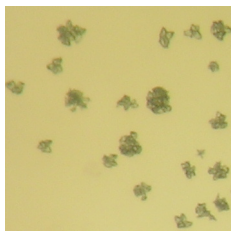
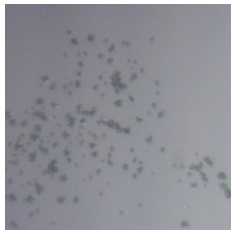

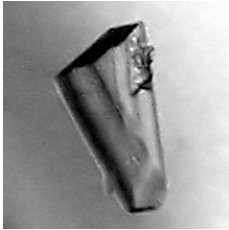
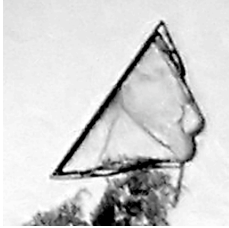
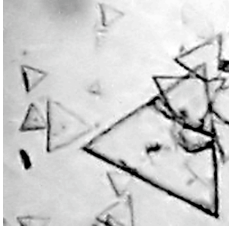
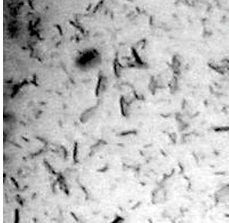
Precipitant	pH	Result	Crystal morphology
30 % MPEG 550 0.1M HEPES	8.2		Birefringent granule
40 % MPEG 550 0.1M HEPES	8.2		Dark granule
36 % PEG 600 0.1M HEPES	7.2		Birefringent cluster
36 % MPEG 2000 0.1M sodium cacodylate	6.5		Birefringent granule
33 % PEG 600 0.2M imidazole malate	5.5		Irregular

Table 32: Initial hits from salt conditions of the Stura Footprint screen

Precipitant	pH	Result	Crystal morphology
0.75–1 M citrate 10mM sodium borate	8.5		Triangular prism
1.32 M $\text{NaH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$	7.0		Triangular prism
1–1.5 M ammonium sulphate 0.15 M sodium citrate	5.5		Triangular plate
2 M ammonium sulphate 0.15 M sodium citrate	5.5		Rod-like plate

6.2.2.2 Optimisation

In order to maximise the number, type and quality of crystals available for further experiments a series of optimisation steps were carried out on each of the initial hit conditions. These involved expanding each set of conditions into a 4×6 screen in which the precipitant concentration and pH were incrementally varied (see Table 33). Only the optimisation of the sodium citrate conditions yielded an improvement in terms of giving large, three-dimensional, single crystals with no visible flaws. These could be further improved by streak seeding into drops containing slightly lower precipitate concentrations than those required for spontaneous nucleation (< 0.75 M citrate).

Precipitant	Buffer	pH
20–35 % MPEG 550	0.1 M HEPES	6.5–9.0
30–45 % MPEG 550	0.1 M HEPES	6.5–9.0
10–40 % PEG 600	0.1 M HEPES	5.5–8.0
10–40 % MPEG 2000	0.1 M sodium cacodylate	5.5–8.0
10–40 % PEG 600	0.2 M imidazole malate	5.5–8.0
0.5–1.5 M citrate	10 mM sodium borate	6.5–9.0
0.5–1.5 M $\text{NaH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$		5.5–8.0
1–2 M ammonium sulphate	0.15 M sodium citrate	5.5–8.0
1.5–2.5 M ammonium sulphate	0.15 M sodium citrate	5.5–8.0

Table 33: Optimisation of ZiCo1 crystallisation conditions. The table shows how the initial hit conditions were expanded to give a set of 4×6 screens in an attempt to optimise the crystallisation of ZiCo1. Only the condition using citrate as a precipitant produced improvements in the quality of crystals obtained.

After six months, fresh drops of ZiCo1 prepared from the original batch of peptide no longer produced crystals. A further batch of peptide was provided courtesy of Beth Bromley, which also failed to produce crystals in any of the conditions tested. Both batches of peptide were analysed by CD spectroscopy and compared with the original data, as shown in Figure 6.4.

The first batch after six months shows no signal, which suggests that the peptide has degraded, thus explaining why no crystals could be produced. The second batch of peptide shows a much weaker signal than the first, which indicates that the peptide is less folded. In addition, the minimum at 209nm is more pronounced than that at 222nm when compared with the original measurement, which is suggestive of an increased contribution of random coil to the CD spectrum. As a result of these observations, further crystallisation attempts were abandoned, and work proceeded with the limited number of crystals obtained.

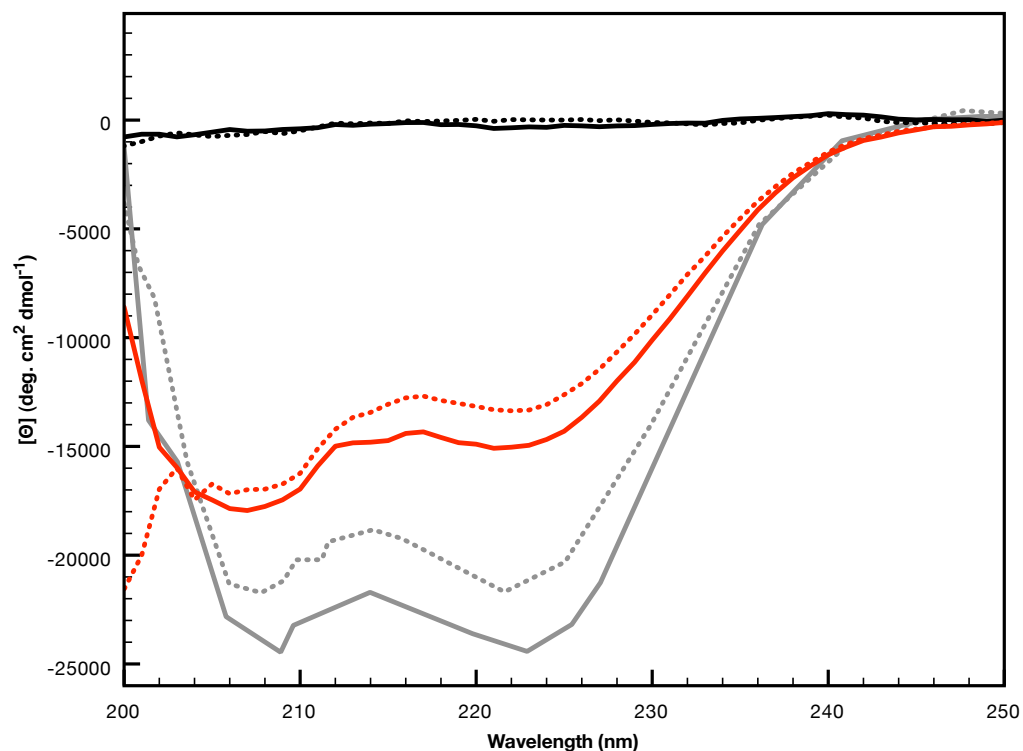


Figure 6.4: Comparison of ZiCo1 batches in 50mM sodium phosphate, pH7.5, 50mM sodium chloride. The CD spectra of 100 μ M ZiCo were recorded in the absence (solid lines) or presence (dotted lines) of equimolar zinc chloride. The first batch after six months is shown in black, while the second batch is shown in red. The original spectrum of the first batch is shown for comparison in grey. The results indicate that there is considerable variability between batches, and that the peptide has degraded over time.

6.3 Data collection and analysis

Following successful crystallisation of the ZiCo1 peptide in 0.75–1M citrate, 10mM sodium borate, pH8.5, data were collected using a Rigaku RU-H3RHB rotating anode X-ray generator equipped with Osmic Max-Flux optics and an RAXIS-IV++ image plate area detector. Cryocrystallography was carried out using the MSC X-Stream cryo system.

6.3.1 Data collection

A single crystal grown in 0.75M sodium citrate, 10mM borate, pH8.5 was collected and briefly immersed in a solution of crystallisation buffer containing 33% glycerol as a cryoprotectant. The crystal was flash frozen at 100 K and a total of 180 images with an oscillation angle of 1° and exposure of 1 min were collected (see Figure 6.5). Upon close inspection of the diffraction images, it was observed that some reflections exhibited split diffraction spots, suggesting disorder in the crystal.

To determine whether this was an inherent feature of the crystals or due to the addition of glycerol as a cryoprotectant, a second crystal grown under the same conditions was mounted at room temperature according to the method of Mac Sweeney and D'Arcy, 2003, and several images were collected with an oscillation angle of 1° and exposure of 10 min. These show well-defined spots to the resolution limit (see Figure 6.6), which indicates that the crystals do not tolerate the addition of glycerol.

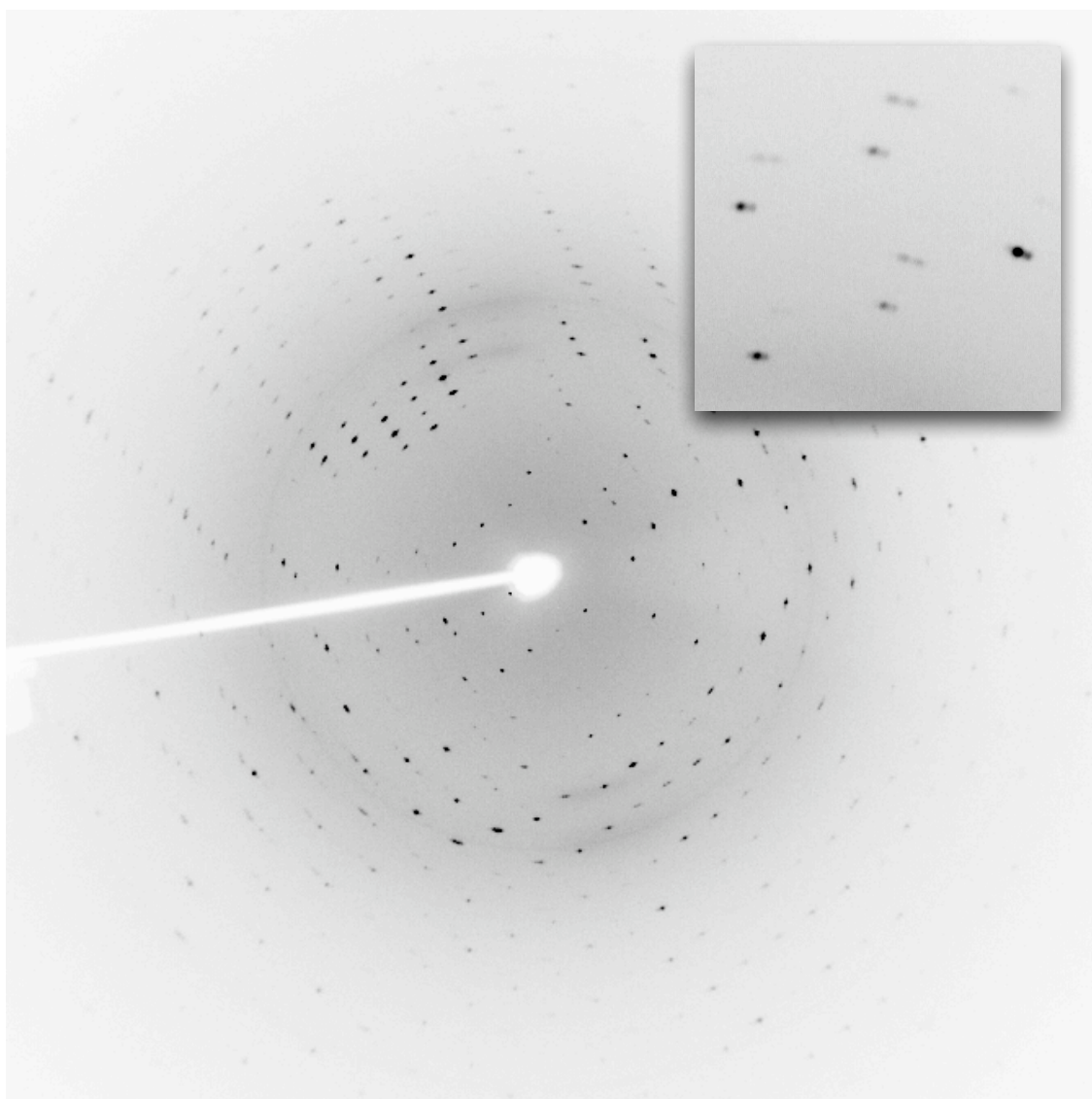


Figure 6.5: Diffraction pattern of ZiCo1 using 33% glycerol as a cryoprotectant. The inset shows pathological spot splitting that was characteristic of crystals treated with cryoprotectants.

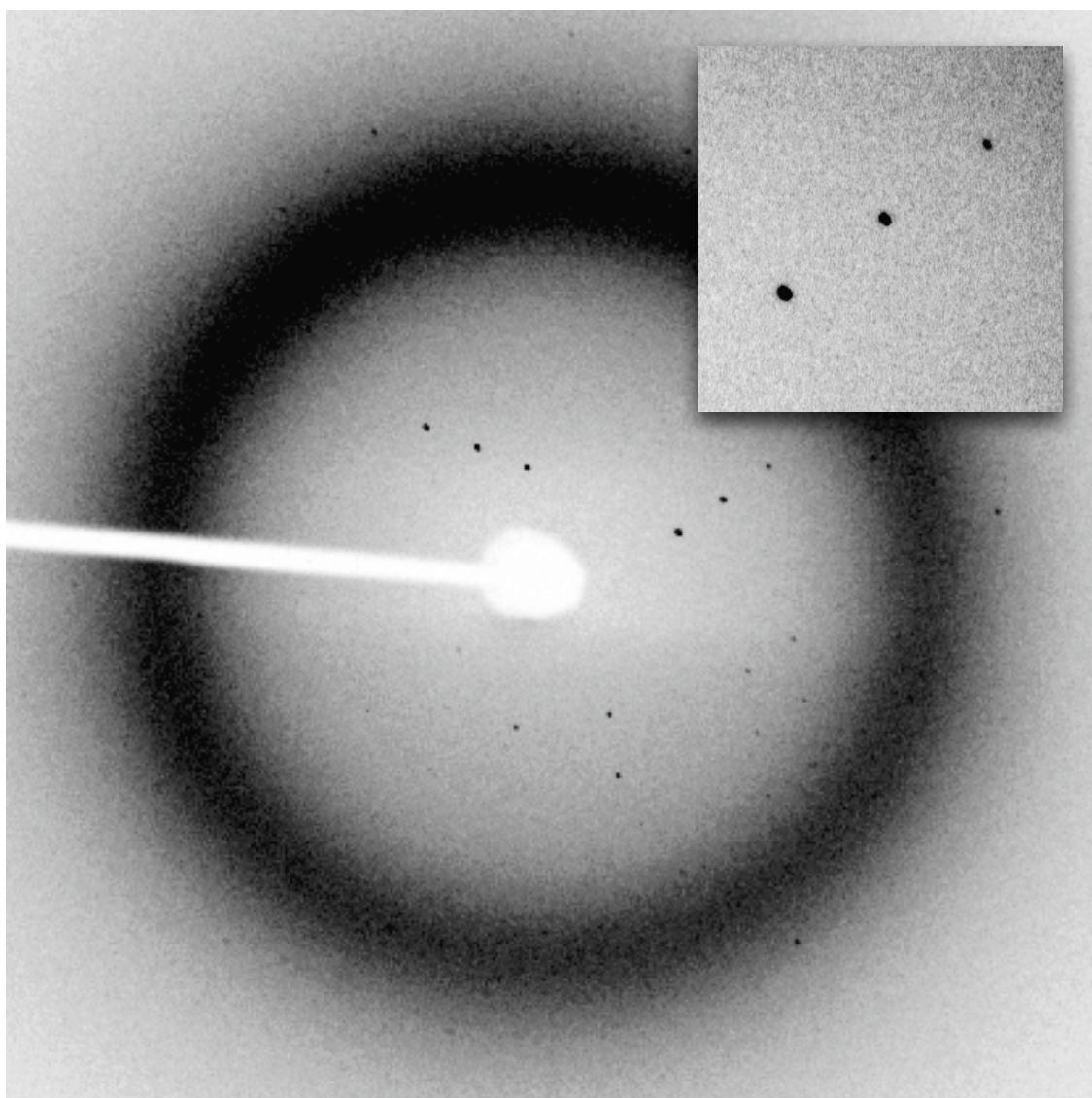


Figure 6.6: Room temperature diffraction of ZiCo1. The inset shows clear, single spots with no obvious pathologies. The diffraction at room temperature was much weaker than at 100 K, resulting in a large background contribution from the solvent, and weaker reflections.

At room temperature all crystals tested diffracted relatively weakly and it was not possible to process the resulting data, ruling out room temperature data collection as an option. Testing with other cryoprotectants such as MPD and PEGs produced similar pathological diffraction to that observed when glycerol was used. Gradually transferring the crystals to the final cryoprotectant concentration in small concentration steps also did not offer any improvement, so an alternative strategy was sought for cryoprotection.

During optimisation of the crystallisation conditions, crystals were successfully grown using higher concentrations of sodium citrate as a precipitant. At 1 M sodium citrate, the mother liquor vitrified upon freezing, so for crystals grown at this concentration no further cryoprotection was necessary. The diffraction images showed clear, single spots and a total of 360° of data were collected with an oscillation angle of 1° and exposure of 1 min (see Figure 6.7).

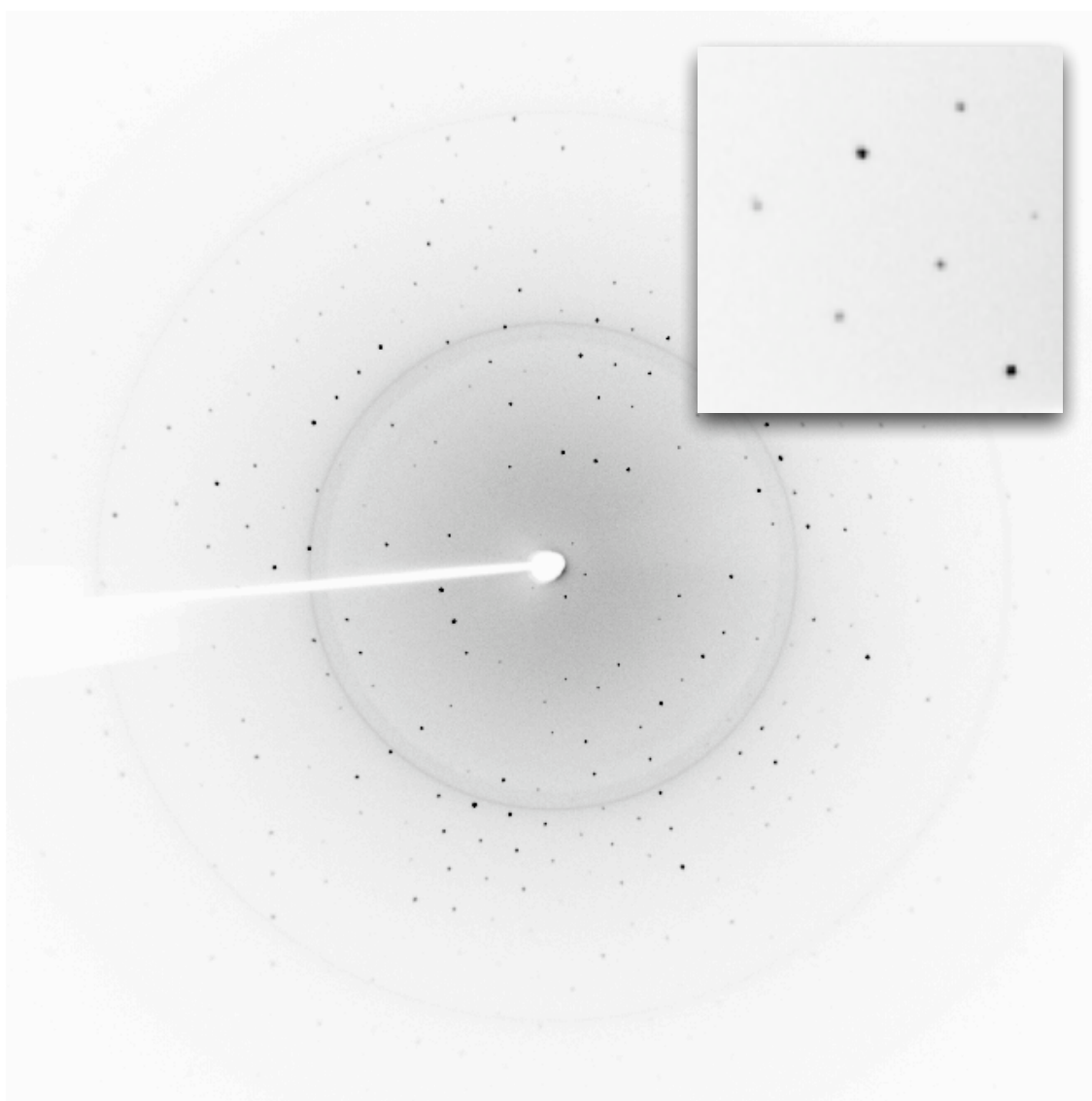


Figure 6.7: Diffraction of ZiCo1 crystals grown in 1M citrate, 10mM borate, pH8.5, without any additional cryoprotectant. The inset shows clear, single spots with no obvious pathologies and a large number of strong reflections. This indicates that high concentrations of citrate act as a suitable cryoprotectant for crystals of ZiCo1 grown under these conditions.

6.3.2 Data processing

The raw data were processed with MOSFLM (Leslie, 1992), which suggested P3 as the most likely space group. The unmerged data were also analysed using Pointless (Evans, 2006), which suggested P312 as the most likely space group. At this stage, it is not possible to assign a space group unambiguously and consequently, the data were processed in various space groups for comparison. Intensities were scaled and merged using Scala from the CCP4 software suite (Collaborative Computational Project, 1994) and the resulting statistics are shown in Table 34. The statistics show that the data process sensibly in all three space groups shown, with processing in P312 giving a slight increase in the value of R_{merge} when compared with P3 (4.9% as against 4.8%).

	P1			P3			P312		
Unit cell parameters	32.0	32.1	57.6	32.1	32.1	57.6			
	90.3	90.1	119.8	90	90	120			
Resolution (Å)	25.0–1.6								
Reflections	71884								
Unique	18543			7661			4342		
Completeness (%)	68			85			90		
Multiplicity	3.6			9.4			16.6		
R_{merge} (%)	7.1			4.8			4.9		
$I/\text{sd}(I)$	14.5			37.6			48.7		

Table 34: Comparison of processing statistics in various space groups. The predicted space group is P312, which gives good completeness and multiplicity and an acceptable R_{merge} value. Processing in P3 also gives similar values, while the statistics in P1 are shown for comparison.

In order to determine the most likely number of molecules in the asymmetric unit for the various space groups, the Matthews coefficient and solvent content were calculated,

along with their associated probabilities for each number of molecules in the asymmetric unit (N_{ASU}). The results are presented in Table 35. Six molecules were given as the most likely number in the unit cell for space groups P3 ($N_{\text{ASU}} = 2$) and P312 ($N_{\text{ASU}} = 1$), while in P1 the most likely number was seven ($N_{\text{ASU}} = 7$). In space group P3, the V_{M} for a single molecule ($5.36 \text{ \AA}^3 \text{ Da}^{-1}$) and the corresponding solvent content (77%) are improbably high ($P(N_{\text{ASU}}) = 0.01$). However, there are reported structures with solvent contents in this region Dias, 2002, so while it is improbable that we have one molecule per asymmetric unit in P3 (or indeed three molecules per asymmetric unit in P1) there is still a small possibility that this may be the case.

Space group	N_{ASU}	$P(N_{\text{ASU}})$ overall	V_{M} ($\text{\AA}^3 \text{ Da}^{-1}$)	V_{s} (%)
P1	6	0.31	2.68	54.10
	7	0.38	2.30	46.45
P3	2	0.79	2.68	54.10
P312	1	1.00	2.68	54.10

Table 35: Matthews' coefficient in various space groups. The most likely number of molecules in the asymmetric unit (N_{ASU}) for each space group is shown, as well as their associated probabilities. The corresponding Matthews coefficient (V_{M}) and solvent content (V_{s}) are also indicated.

6.3.3 Twinning tests

Crystal twinning is a growth disorder that occurs when two or more crystals of the same species intergrow in a way that results in good molecular packing but violates crystal symmetry (Yeates and Fam, 1999). Twinned crystals are thus composed of multiple domains whose orientations are related in some way. While some twinned crystals are instantly recognisable by their shape, many appear to be single crystals, and the presence of twinning can only be detected from the diffraction data.

6.3.3.1 *Types of twinning*

If the crystal lattices of individual domains are superimposable in only two dimensions, the result is known as ‘non-merohedral’ or ‘epitaxial’ twinning. This is easily recognised from the diffraction pattern, which will display multiple interpenetrating lattices. However, if the lattices of the twin domains overlap in all three dimensions the situation becomes more complicated. This type of twinning is known as ‘merohedral’ twinning and manifests itself in crystal systems where the unit cell parameters allow the possibility of higher symmetry than the true symmetry of the crystal. This is generally the case with trigonal, tetragonal, hexagonal and cubic systems, although monoclinic, triclinic and orthorhombic systems may also exhibit twinning if the unit cell parameters allow it. Merohedrally twinned macromolecular crystals are almost always composed of just two distinct domains related by a twofold symmetry operator, and this special case is termed ‘hemihedral’ twinning.

6.3.3.2 *Implications for structure solution*

While the diffraction pattern of merohedral twins may look normal, owing to the superimposability of the lattices, each recorded intensity actually contains contributions from the two twin-related reflections in a ratio that is dependant on the volume occupied by each twin domain. As a result, the diffraction pattern acquires varying degrees of additional symmetry as imposed by the twin law. If one domain is much larger than the other, then the crystal is partially twinned. In this case, the true crystallographic intensities can be recovered mathematically at the expense of magnifying the measurement errors by a factor of roughly $1/(1-2\alpha)$, where α is the twin fraction.

However, as the volumes of the twin domains approach equality and α approaches 0.5, the measurement errors become increasingly large, and so for a perfectly twinned crystal it is impossible to mathematically de-twin the data. The diffraction pattern will display erroneously high symmetry, and if twinning is not detected, the true space group may never be established. If the twinning is detected, such a case may be solved either by searching for new crystals with partial or no twinning, or by attempting to determine the structure without knowing the true crystallographic intensities. This is possible in practice by using molecular replacement.

6.3.3.3 *Detecting merohedral twinning*

There are numerous indicators of twinning, several of which were detected in the ZiCo1 data. Firstly, the crystal system appears to be trigonal, in which twinning is common. Secondly, the R_{merge} for the higher symmetry space group P312 is only slightly greater than that for P3, which is characteristic of twinned data (CCP4 Wiki). In the case that the data are twinned, the most likely scenario is that the true space group is P3, with a twinning operator that mimics the symmetry of P312. Consequently, the P3 data were subjected to various twinning tests.

As mentioned previously, in the case of hemihedral twinning, each measured reflection will have contributions from both twin domains. It is unlikely that these will both have extremely high or extremely low intensity values, so the distribution of intensities in a twinned data set will follow a sigmoidal distribution, where there are fewer extreme values than expected. Indeed, the intensity distributions of the ZiCo1 data set as determined by TRUNCATE (French and Wilson, 1978) indicate that the crystal is twinned (see Figure 6.8). Furthermore, plots of the k^{th} moments of E , another indicator of twinning, are close to the predicted values for a perfectly twinned crystal (see Figure 6.9).

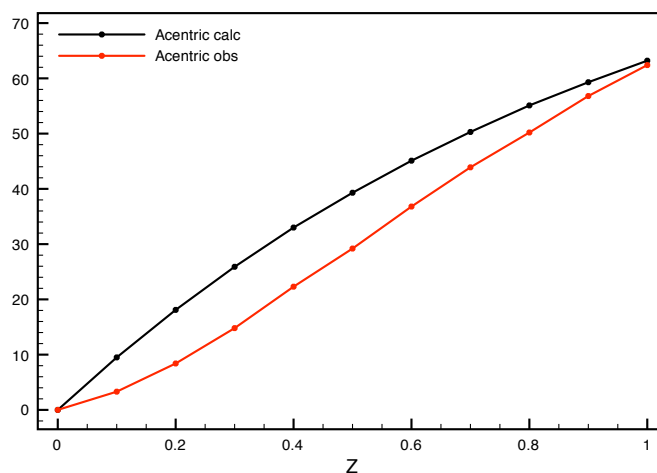


Figure 6.8: Intensity distribution of P3 data. The observed intensity distribution of acentric reflections is sigmoidal, and does not correspond to the calculated theoretical distribution. This is a known indicator of twinning.

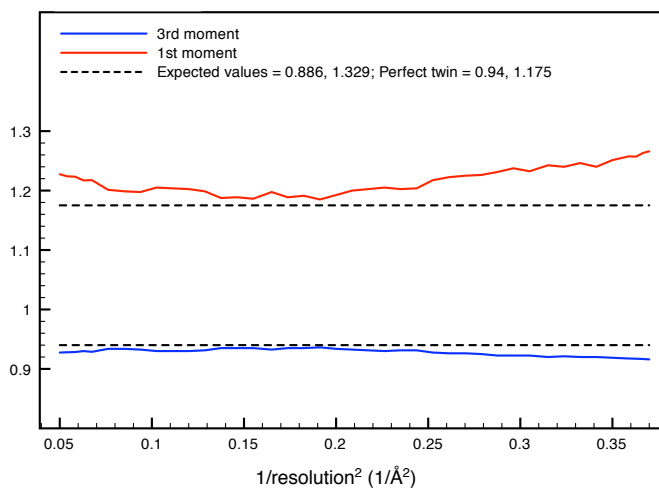


Figure 6.9: Plot of the k^{th} moments of I for ZiCo data processed in space group P3. The expected values for perfectly twinned data are shown (dashed lines). The moments lie close to the values expected for a perfectly twinned crystal.

6.3.3.4 *SFCHECK*

Further twinning tests were performed using *SFCHECK*, which determined the twinning operator to be $(h+k, -k, -l)$ with a twin fraction of 0.471 in space group P1 (see Figure 6.10). In P3, the twinning operator was given as $(-h, h+k, -l)$ with a twinning fraction of close to 0.5 (see Figure 6.11). In space group P312, the data appeared to be twinned by $(-h, -k, l)$ with a twin fraction of 0.2. This suggests that the crystal is almost certainly twinned, and de-twinning the data in P3 or P1 would be impossible. Several other crystals were tested and all exhibited a similar degree of twinning, so the decision was made to proceed with the available data and to attempt a solution via molecular replacement.

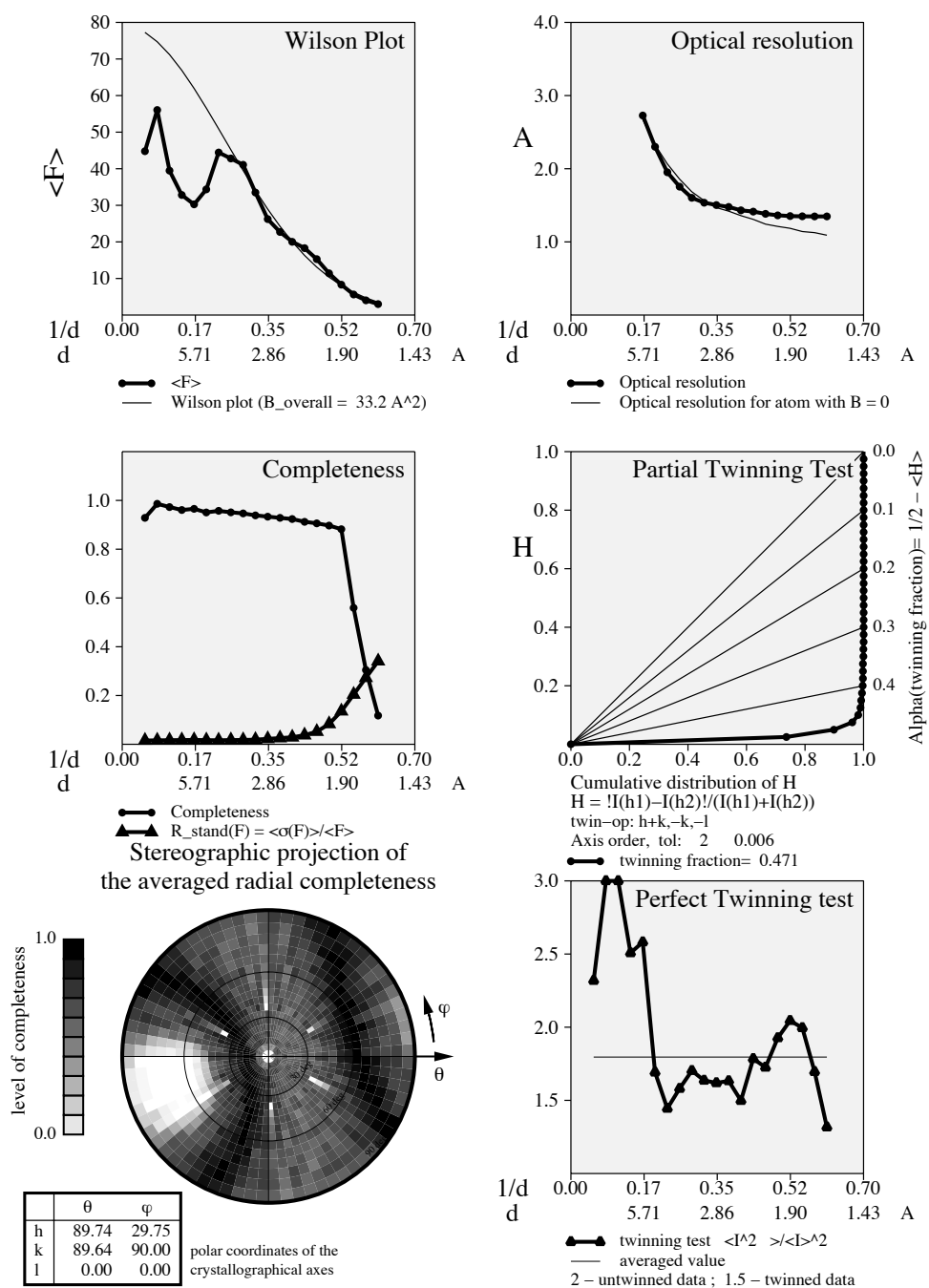


Figure 6.10: SFCHECK output for the ZiCo data processed in P1. The Wilson plot shows some deviation between 3–2 Å, and the data are relatively complete. The partial twinning test gives a twinning fraction of 0.471 and the perfect twinning test shows that the data is perfectly twinned over part of the resolution range.

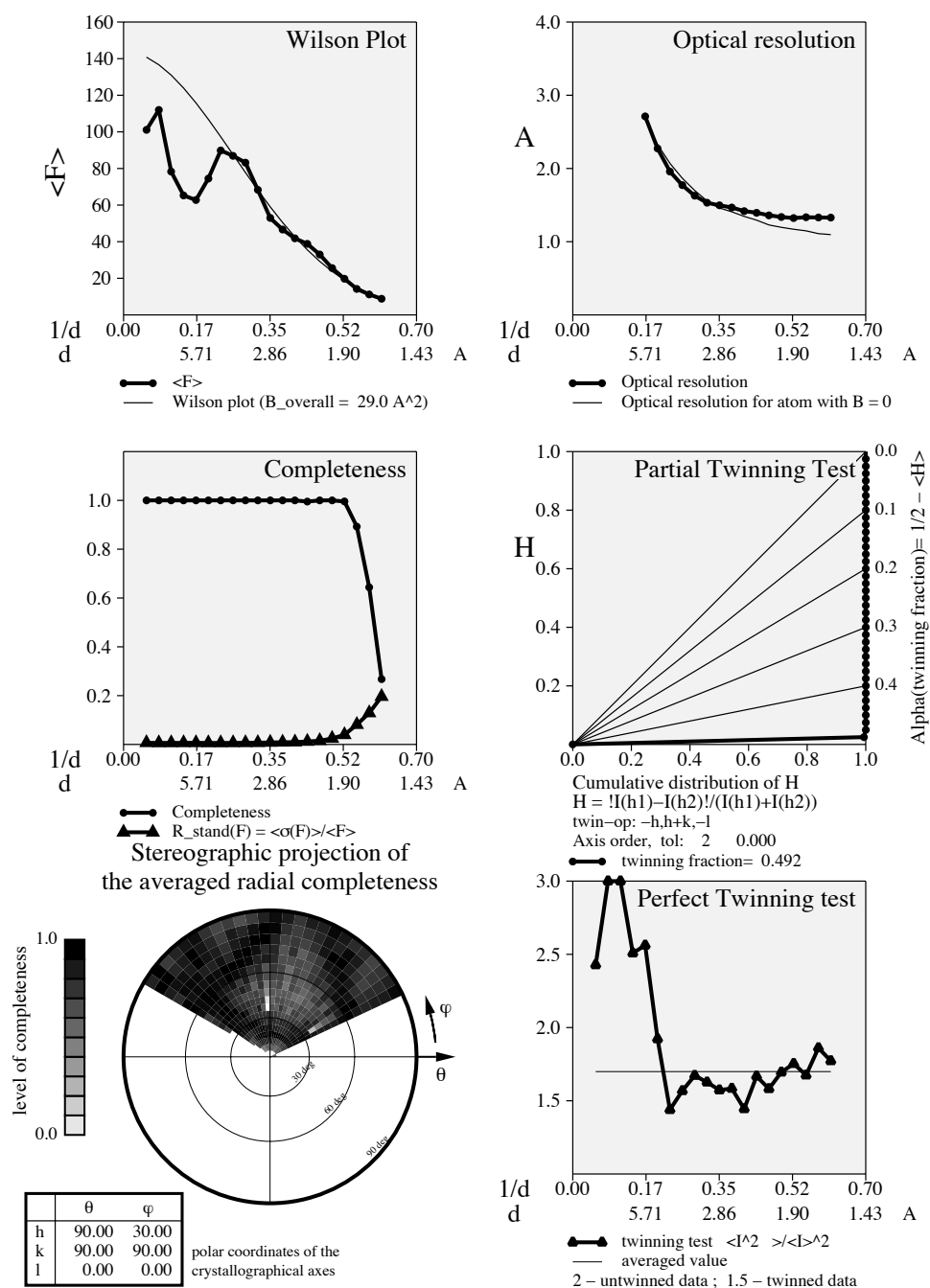


Figure 6.11: SFCHECK output for the ZiCo data processed in P3. The Wilson plot shows some deviation between 3–2 Å, and the data are complete to around 2 Å. The partial twinning test gives a twinning fraction of 0.492 and the perfect twinning test shows that the data is perfectly twinned over part of the resolution range.

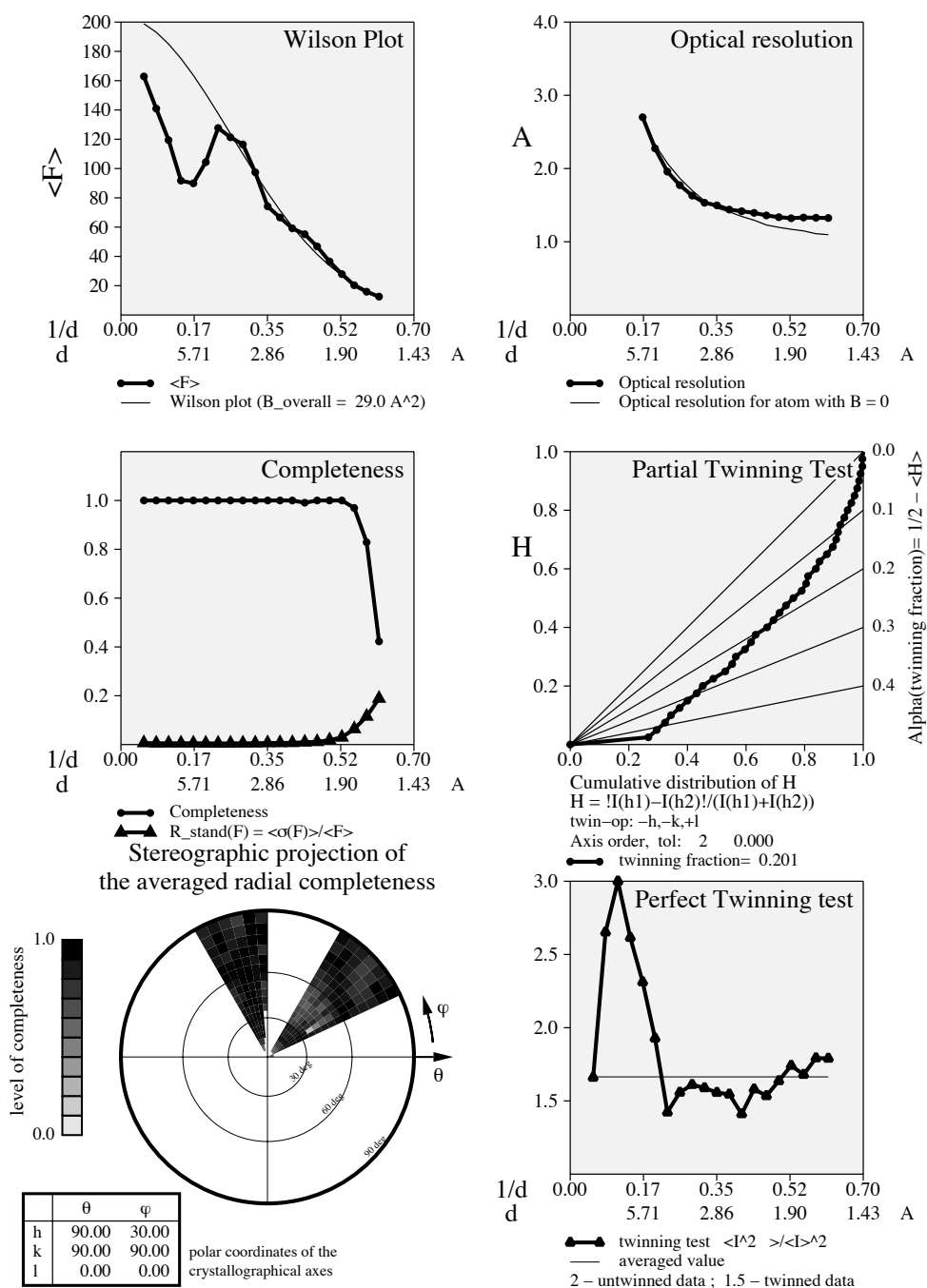


Figure 6.12: SFCHECK output for the ZiCo data processed in P312. The Wilson plot shows some deviation between 3–2 Å, and the data are complete to around 2 Å. The partial twinning test gives a twinning fraction of 0.201 and the perfect twinning test shows that the data is perfectly twinned over part of the resolution range.

6.4 Molecular replacement

Molecular replacement was chosen as an initial approach to obtain phases that would allow the crystal structure of ZiCo1 to be solved, since it can be applied in the case of twinned crystals. In order for molecular replacement to be successful it requires a suitable model that is sufficiently similar to the unknown structure to provide initial estimates of the phases to allow a structure solution to be found. The traditional approach to selecting a model involves searching for known structures that bear a significant sequence similarity to the unknown protein structure to be solved. However, because of the novel nature of the ZiCo sequence this approach was not suitable and models were instead chosen based on the principles of the ZiCo design, in the hope that the crystallised structure would be sufficiently similar to the proposed structure to allow a solution to be found.

6.4.1 Zinc finger model

ZiCo1 crystallised in the presence of zinc, so the initial working hypothesis was that the peptide had crystallised as a zinc finger. The PDB was initially queried for C₂H₂ zinc fingers containing the motif X₂CX₂C₁₂H₅H₂ that characterises the ZiCo1 sequence. It was assumed that the peptide would fold in the same manner as naturally occurring zinc fingers that feature cysteines in place of the histidine residues substituted in the designed ZiCo sequence. This search revealed no known structures featuring the motif, so the decision was made to proceed with an arbitrarily chosen zinc finger structure, 1ZNF, which was determined by NMR.

A model was constructed by selecting a single structure from the ensemble contained within the PDB file, truncating all residues to alanine and removing disordered residues from the N and C termini. An extensive search using a number of

different molecular replacement programs ensued, but no solutions were found. Consequently, it was posited that ZiCo1 differed greatly from the model, or that it may not have crystallised as a zinc finger.

The next step undertaken was to determine whether the unknown structure did in fact contain zinc, which could potentially be used to obtain phase information using anomalous scattering techniques. Zinc atoms are much heavier than the other contents of the unit cell and should therefore scatter incident X-ray radiation more strongly. Strongly scattering atoms can be located by searching for large non-origin peaks in the native Patterson map. A native Patterson was calculated using the program FFT (Ten Eyck, 1973) and PEAKMAX was used to search for peaks. The output is given below in Table 36, with peaks listed in order of their normalised heights.

Table 36: Native Patterson peak list. The first peak is the origin peak.

Order	No.	Site	Height/Rms	Grid			Fractional coordinates			Orthogonal coordinates		
1	1	1	86.31	0	0	0	0.0000	0.0000	0.0000	0.00	0.00	0.00
2	7	4	8.18	59	0	10	0.9833	0.0000	0.0922	31.50	-0.03	5.31
3	13	4	8.18	1	0	100	0.0106	0.0000	0.9078	0.25	-0.33	52.29
4	14	4	6.54	59	0	100	0.9833	0.0000	0.9062	31.42	-0.33	52.20
5	8	5	5.53	0	4	19	0.0000	0.0635	0.1741	-1.03	1.71	10.03
6	11	7	5.04	4	4	91	0.0622	0.0668	0.8255	0.84	1.57	47.55
7	3	2	4.81	0	6	6	0.0000	0.0995	0.0565	-1.59	2.76	3.26
8	12	6	4.75	56	0	91	0.9373	0.0000	0.8252	29.95	-0.30	47.54
9	15	3	4.06	54	1	105	0.9025	0.0139	0.9521	28.60	0.04	54.85

The first peak in Table 36 is the origin peak, and all other peaks are much less prominent ($< 15\%$ of origin) and have fractional co-ordinates close to the origin, so can be assumed to be shoulders of a broad origin peak. No other significant non-origin peaks were found, indicating that zinc was not present in the crystal and hence implying that ZiCo1 had not crystallised as a zinc finger.

6.4.2 Single helix models

Since the crystals did not appear to contain zinc, it was concluded that ZiCo1 had therefore not crystallised as a zinc finger, despite the presence of zinc in the crystallisation conditions. An alternative hypothesis was that ZiCo1 may have instead crystallised in its trimeric, alpha-helical form, so to test this idea the molecular replacement searches were repeated with a helical model.

The PDB entry 1GCM is a parallel, trimeric coiled coil variant of the GCN4 leucine-zipper dimerisation domain, which was selected as a suitable model for the proposed ZiCo1 structure. A sequence alignment shows key similarities and differences between the two sequences (see Figure 6.13). An initial series of models was prepared by selecting a single helix from the 1GCM structure and truncating all residues to alanine. Chains of various lengths from 9 to 27 residues were generated by removing a number of residues from the N terminus.

```

1GCM                XRMKQIEDKIEEILSKIYHIENEIARIKKLIGER 34
ZICO1               -YIHALHR---KAFAKIARLERHIRALEHAA--- 27
                   :: :.      : ::** ::*..*   :::

```

Figure 6.13: Sequence alignment of ZiCo1 and 1GCM. Conserved residues are marked with an asterisk (*), conserved substitutions are marked with a colon (:), and semi-conserved substitutions are marked with a period (.).

Molecular replacement was attempted in AMoRe, Molrep and PHASER (McCoy et al., 2007) using data indexed in P3. The initial searches were unsuccessful, with AMoRe and Molrep failing to find any solutions at all, while those found by PHASER appeared unlikely to be correct. The best solution from PHASER was using a 24 residue polyalanine model and searching for two copies in the asymmetric unit. The Z-scores for the rotation function and translation function for each copy were $RFZ = 3.7$;

TFZ = 4.7 and RFZ = 2.8; TFZ = 6.0, while correct solutions typically have Z-scores > 5 and > 8 for the rotation and translation functions respectively. However, inspection of the maps in Coot (Emsley and Cowtan, 2004) revealed the presence of additional features in the maps suggestive of side chains, suggesting that the solution may be correct despite the low scores (see Figure 6.14).

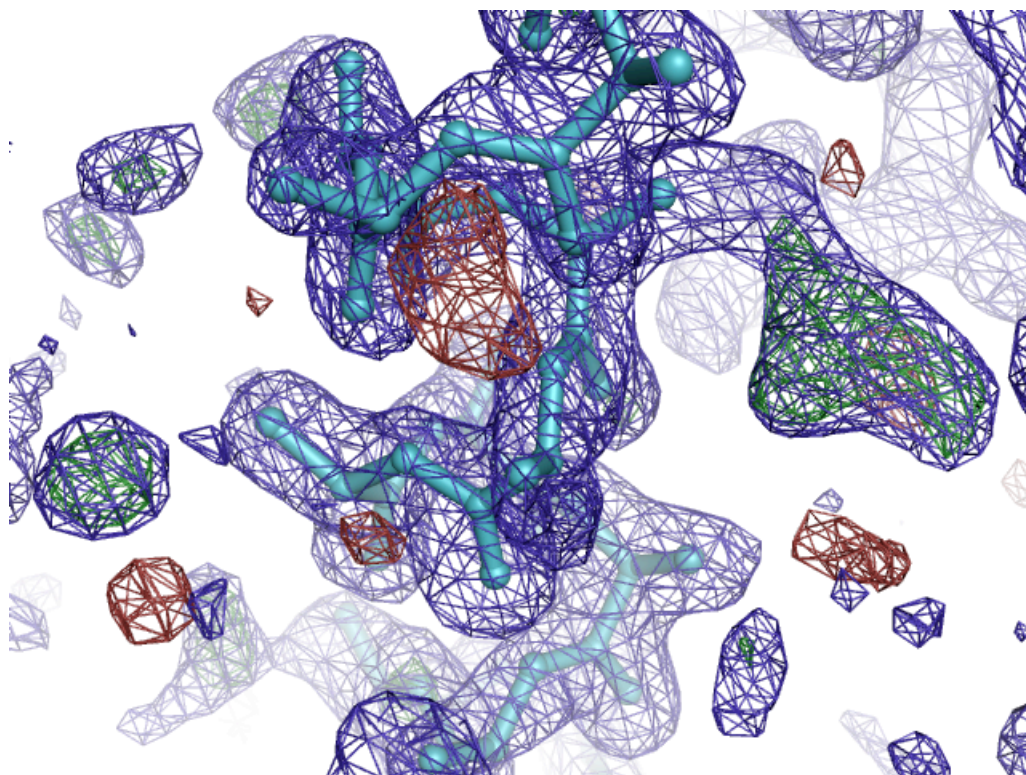


Figure 6.14: Detail of MR solution using a 24 residue polyaniline search model in P3. The calculation was carried out in space group P3, searching for two copies per asymmetric unit. Maps are contoured at 1.5σ , with electron density (blue) and an $mF_O - DF_{\text{calc}}$ difference map showing regions of positive (green) and negative (red) density. The positive density (right) is indicative of a side chain.

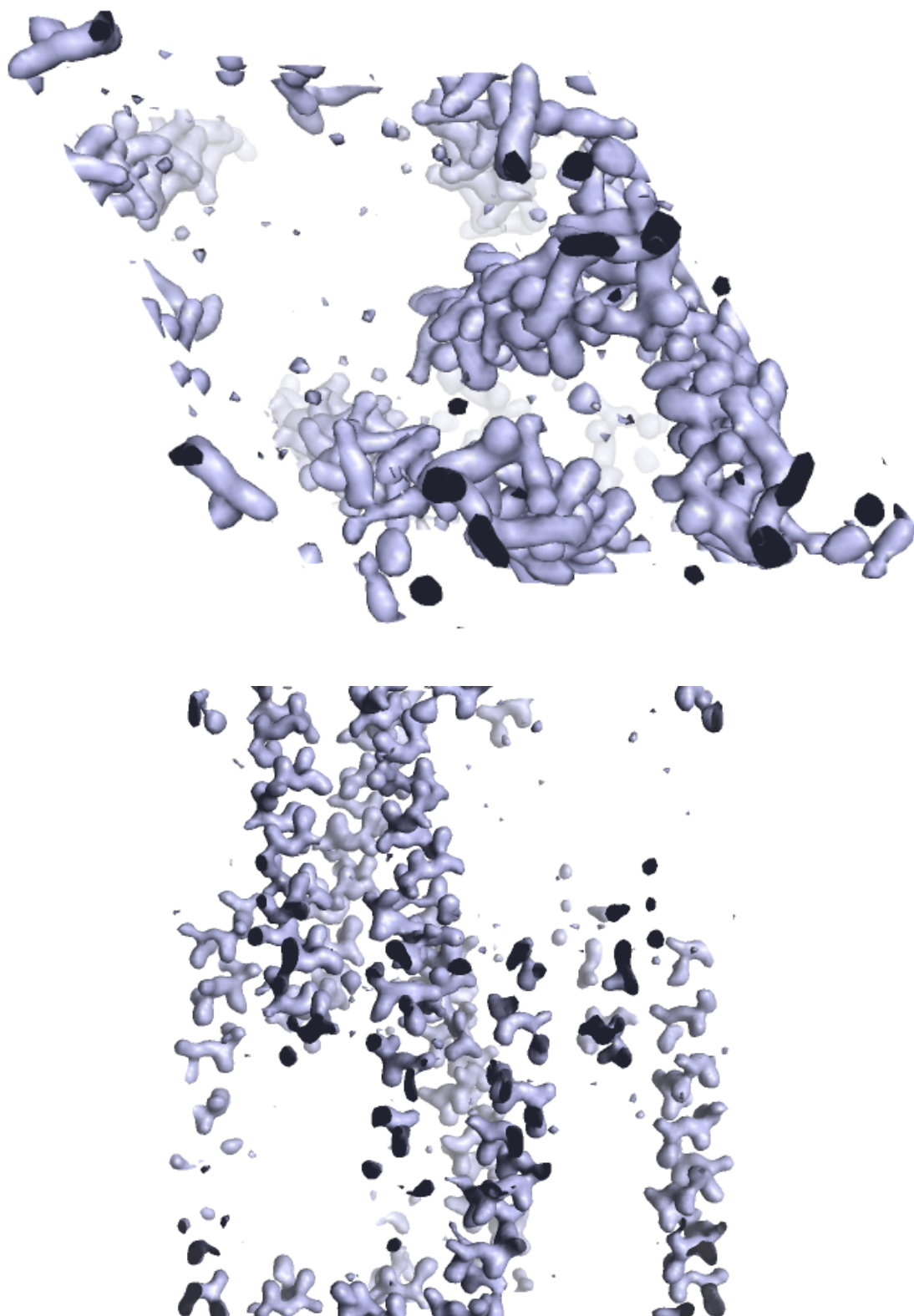


Figure 6.15: Packing of MR solution using a 24 residue polyaniline search model. The calculation was carried out in space group P3, searching for two copies per asymmetric unit. Maps are shown contoured at 2σ , viewed along the coiled coil axis (top) and from the side (bottom). The three helices that form the coiled coils are related by crystallographic symmetry.

The linear arrangement of the polypeptide chain in a coiled coil structure may explain why no single solution was found in the MR attempts using a coiled coil model. Since the Patterson map contains both self-vectors and cross-vectors between symmetry-related molecules, the search is performed within a defined radius from the origin of the Patterson function. This works well for globular molecules, but in the case of elongated molecules (such as a single helix) the intramolecular vectors may be much longer than the intermolecular vectors. Reducing the radius of the search may improve the results to a point, although no significant improvements in the outcome of the MR procedure were found when this was attempted.

In terms of maximum likelihood, the low scores might again be explained by the unusual structure of a single helix model. The translation function is generally applied after the rotation function, to pick out the correct position of the molecule within the unit cell. However, a single polyalanine helix translated to various positions along its axis may give a set of highly similar scores for the translation function, meaning that no single solution stands out. Finally, the solution of the phase problem for the ZiCo1 data by MR using a single polyalanine helix as a model might be hampered by the presence of twinning.

On inspection of the maps, the order of side chains could not be determined, so manual model building attempts using the results of these MR searches were not pursued further. The results of molecular replacement are biased towards the input model used in the search. This is because the phase information, which is taken from the model structure, contains more structural information than the experimentally recorded intensities. The more similar the model is to the actual structure, the less biased the resulting electron density maps will be, and so one way of producing better maps is to improve the model. To try to improve, the quality of the ZiCo1 maps a series of modified MR models were constructed.

6.4.3 Key residue models

Polyalanine models are useful when there are large differences between the sequences of the model and target structures or when the position of side chains is uncertain. In the case of ZiCo1, the structure is novel and no suitable sequence model exists, so adding side chains to the model means that there will be many more atoms that are more likely to be incorrectly positioned, thereby making it a worse model. However, estimating which side chains are likely to be in a fixed position and only adding these to the model may offer an improvement. Side chains likely to be in a fixed position are those involved in strong interactions: in this case the hydrophobic and electrostatic interactions at the coiled coil interfaces. In addition, aromatic side chains may offer recognisable features in the electron density. Furthermore, the central portion of the helix is likely to be more folded, as coiled coils have a tendency to fray at the ends (Prof. Woolfson, personal communication). The sequence in Figure 6.16 shows key residues highlighted in the ZiCo1 sequence.

1 1 0 2 0
 Y I H A L H R K A F A K I A R L E R H I R A L E H A A

Figure 6.16: ZiCo1 sequence with key residues highlighted. Hydrophobic (red), basic (green) and acidic (purple) residues are shown. Another key residue is the phenylalanine (F) at position 10. This has a characteristic aromatic side chain, which may offer a recognisable feature in the electron density. In addition, the residue is located in the central region of the sequence, which is likely to be folded and in a fixed position.

A particularly interesting residue is the phenylalanine at position 10 (Phe10). It possesses a benzene ring functional group, which has a very characteristic shape, and is located in the central third of the sequence. There is a four residue insert in the heptad repeat at position 9. This is designed to introduce a local unwinding of the helix and corresponds to the turn linking the α -helix and β -sheet regions of a classical zinc finger

(see Figure 6.17). While this may destabilise the region of the structure near Phe10, the residue is likely to be located on the outside of the coiled coil, due to its bulky size and its position in the heptad repeat (which corresponds to position b in the register). For a discussion of coiled coil sequences, please refer to section 1.6.3.



Figure 6.17: Alignment of the ZiCo1 sequence with the sequence of a classical zinc finger. The sequence register is shown on the first line, with the insertion in the coiled coil heptad repeat highlighted in blue. An aligned classical zinc finger sequence (PDB ID: 1ARD) shows the position of the β -hairpin (yellow) and the α -helix (magenta). The position of the stutter in the coiled coil sequence corresponds to the loop linking the β -hairpin with the α -helix.

The Phe10 model was constructed from a 7-residue polyalanine helix by mutating one of the central residues to phenylalanine using Coot (Emsley and Cowtan, 2004). This procedure places a side chain in the most likely conformation according to a library of information about bond lengths and angles in protein structures. The model was used to search for MR solutions in space group P3, where two molecules are predicted to occupy the asymmetric unit. However, PHASER failed to find a second copy due to steric clashes.

Seven residues correspond to two turns of an alpha helix, so the second solution may be overlapping the first, for example by being placed a turn above or below in the helix. While the top scores for a single molecule were low (RFZ = 3.12; TFZ = 3.06), this can be explained by the incompleteness of the model, which may be only around 10–25% of the scattering contents of the unit cell. The results of a search for a single

copy of the Phe10 model in P3 are shown in Figure 6.14. Despite the low scores, clear helical density can be seen above and below the model, including features that resemble additional side chains.

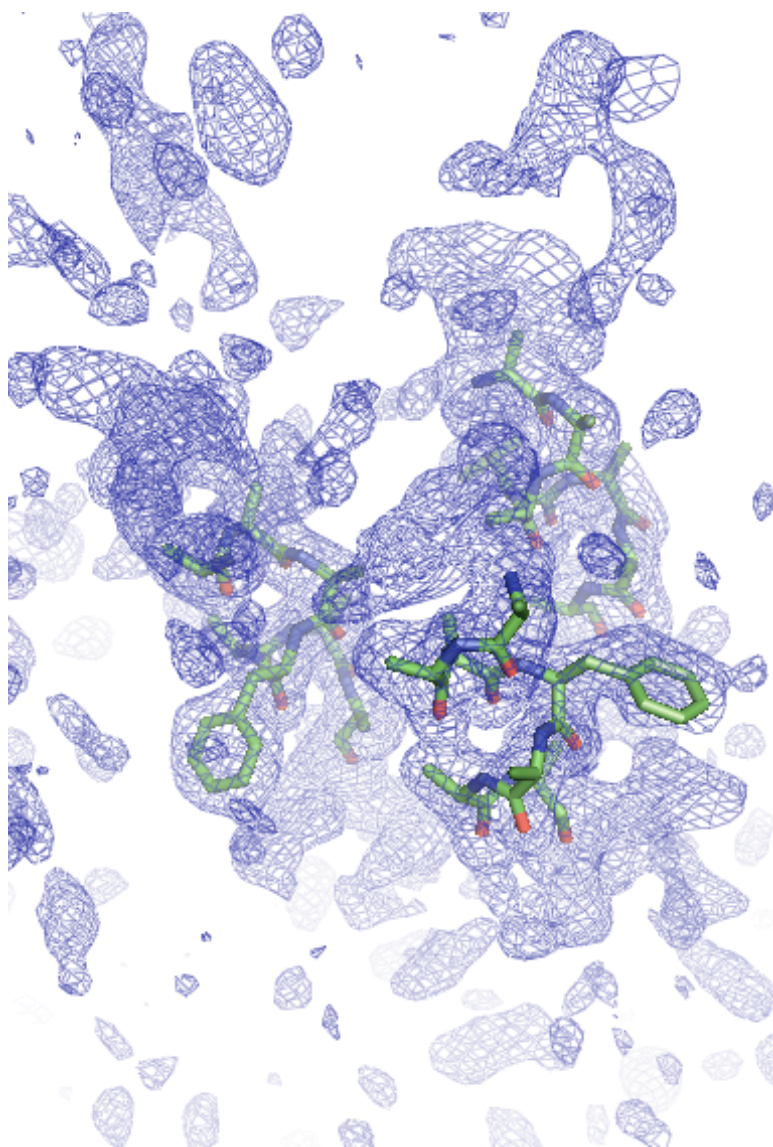


Figure 6.18: MR solution using a 7-residue Phe10 model in P3. The map is contoured at 1.5σ and shows clear helical density above and below the model. Additional features indicative of side chains can be seen below the phenylalanine ring.

6.4.4 Trimeric coiled coil models

The molecular replacement using PHASER was repeated, using a 24 residue trimeric coiled coil as a search model. The model was again generated by truncating the side chains of PDB entry 1GCM to alanine. The search was performed using data indexed in P1 and allowed a search for two copies of the trimer.

Several similar solutions were found and the best result had relatively high scores of $\text{RFZ} = 5.3$; $\text{TFZ} = 100.0$ and $\text{RFZ} = 5.9$; $\text{TFZ} = 7.5$ for the two respective copies. The likelihood log gain (LLG) increases from 168 for the first copy, to 362 for the second. This suggests that these two copies of the trimer describe the observed data more accurately than a single trimer. The corresponding electron density maps are shown in Figure 6.19. These show the close packing of the coiled coil trimers within the crystal structure with this solution.

While the MR attempt using a trimeric coiled coil model in space group P1 appears to have been successful ($\text{TFZ} = 7.5$), it was still not possible to model the resulting density manually. In an attempt to dock the sequence of the ZiCo1 peptide onto the models obtained by MR, these were used as input for an automated model building procedure, as described in the next section.

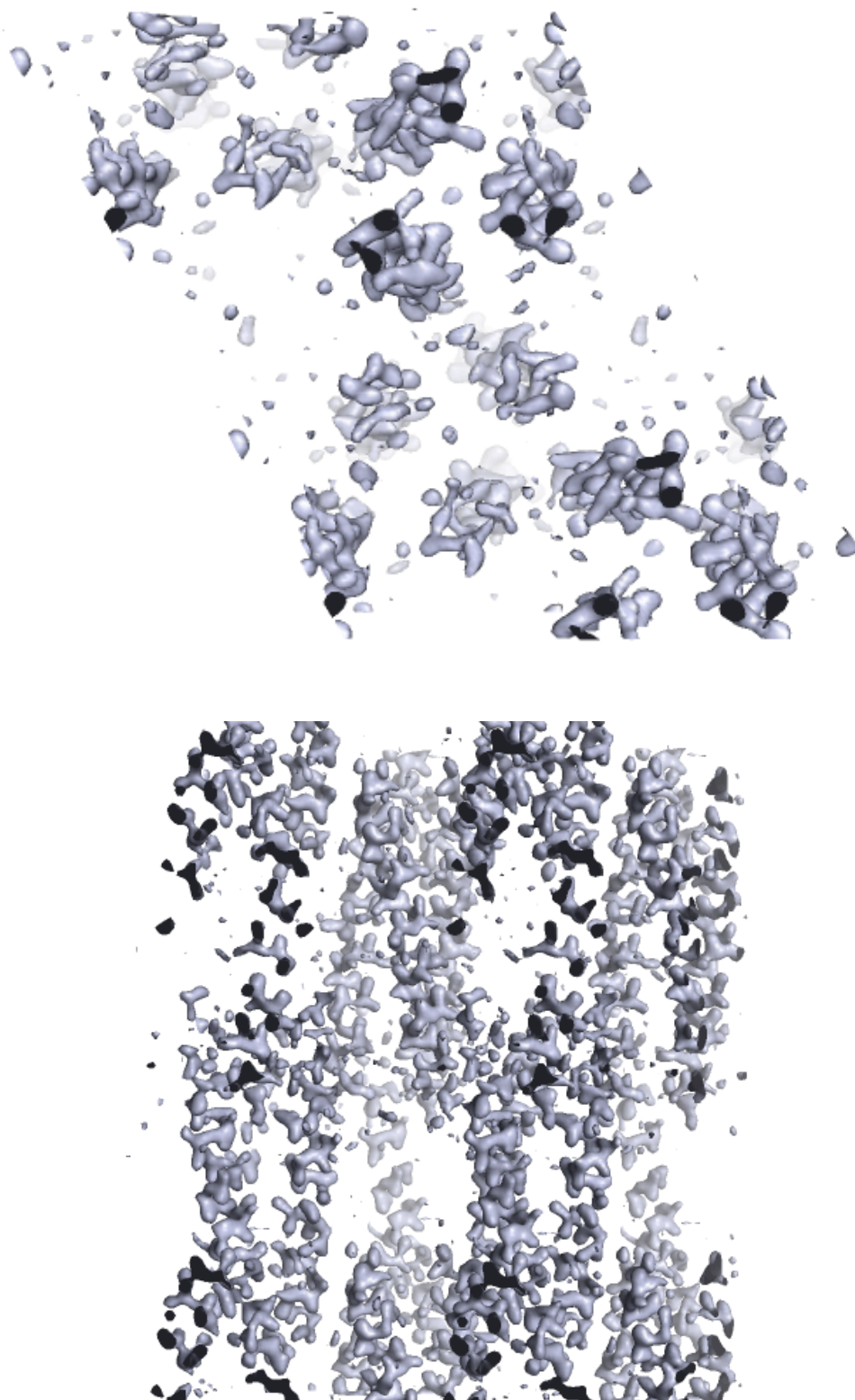


Figure 6.19: MR solution for a 24-residue polyaniline trimeric coiled coil model in P1. The electron density is shown looking down the coiled coil axis (top) and from the side (bottom). This solution shows good packing of the molecules, with six helices in the unit cell.

6.5 Automated model building and refinement

ARP/wARP is a software package that combines electron density interpretation with maximum likelihood refinement in REFMAC to perform automated model building including automated protein main chain tracing, sequence docking, ligand and solvent building (Perrakis et al., 2001). The results of successful MR searches were used as a starting point for this procedure, which was run using default settings. The best solution was using a single helix model in space group P3; the results are illustrated in Figure 6.20. Two chains with 49 residues were docked (99% complete) with an R-factor of 0.35. Further refinement using REFMAC gave an R_{free} of 0.28, while the R-factor decreased to 0.23.

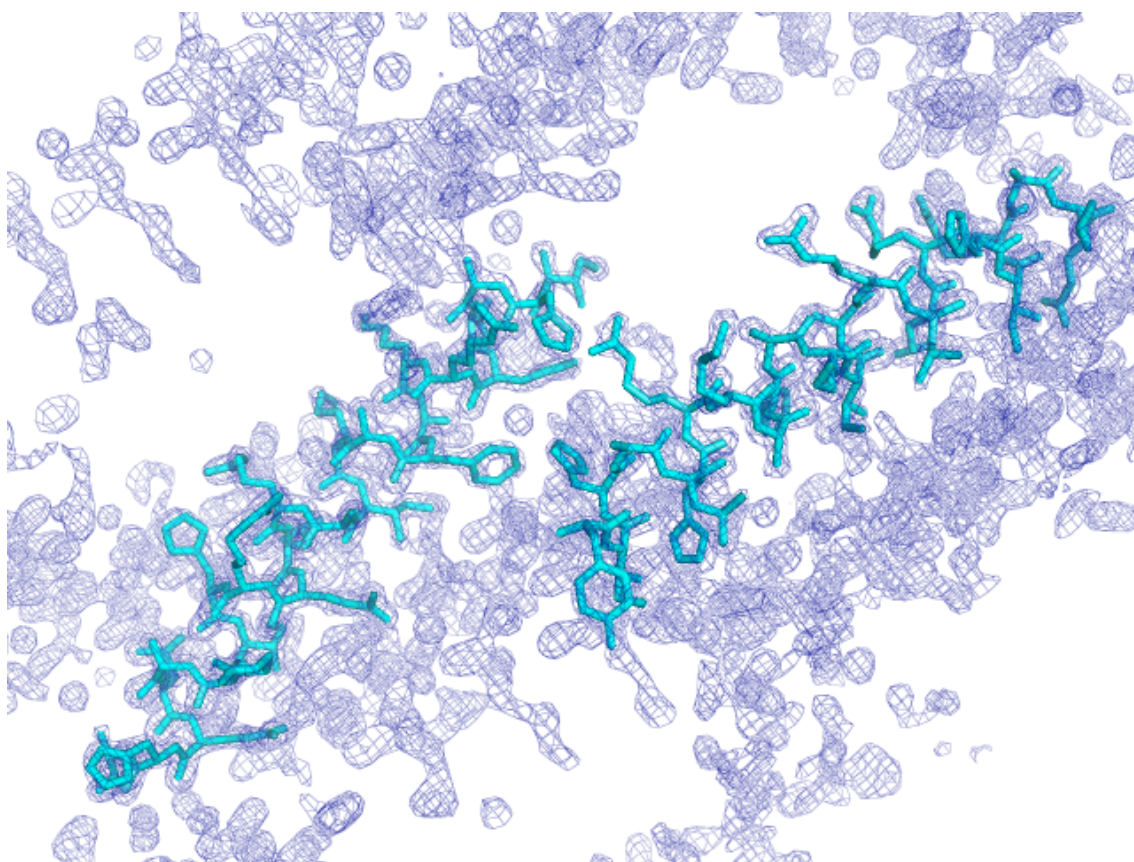


Figure 6.20: Model of ZlCo1 in P3 built using ARP/wARP. Chains: 2; residues docked: 49; estimated correctness of the model: 99 %. Sequence coverage is 98 %; R-factor = 0.35.

6.6 *Ab initio* phasing

While the refined model appears to be correct, although not complete, the phases are likely to be heavily biased towards the initial model used in molecular replacement. One way of obtaining unbiased phases is by using direct methods. These are limited to small molecules and high-resolution data, but may just be applicable to ZiCo1. The peptide contains a predicted 469 atoms (including hydrogens) and the maximum resolution of the data is around 1.6 Å.

The CCP4 program ACORN (Foadi et al., 2000) can be used to calculate the electron density for a small molecule starting from random positioning of atoms. It requires atomic resolution data (better than around 1.2 Å) and can solve structures up to several thousand atoms. To solve the protein phases *ab initio*, a substructure composed of a small fraction of the scattering matter of the unit cell is positioned. This is used to generate a starting set of phases that are slightly better than random, and the full structure is then developed from this phase set. The positions of the atoms are refined and the phases are improved using density modification procedures over the course of several cycles.

The maps calculated using ACORN for ZiCo are poor but unbiased (see Figure 6.21). They clearly show regions of density corresponding to approximately 3 turns of a helix, arranged in a trimer. The helical axes are parallel to the crystallographic c axis, and the ends of the helices are poorly defined. Individual polyalanine helices can be positioned in the density, as well as a trimeric coiled coil. However, the maps were too poor for automated model building and could not be interpreted well enough to begin manual model building. However, they do clearly demonstrate the trimeric, helical nature of the ZiCo1 crystal structure and corroborate the results obtained by molecular replacement.

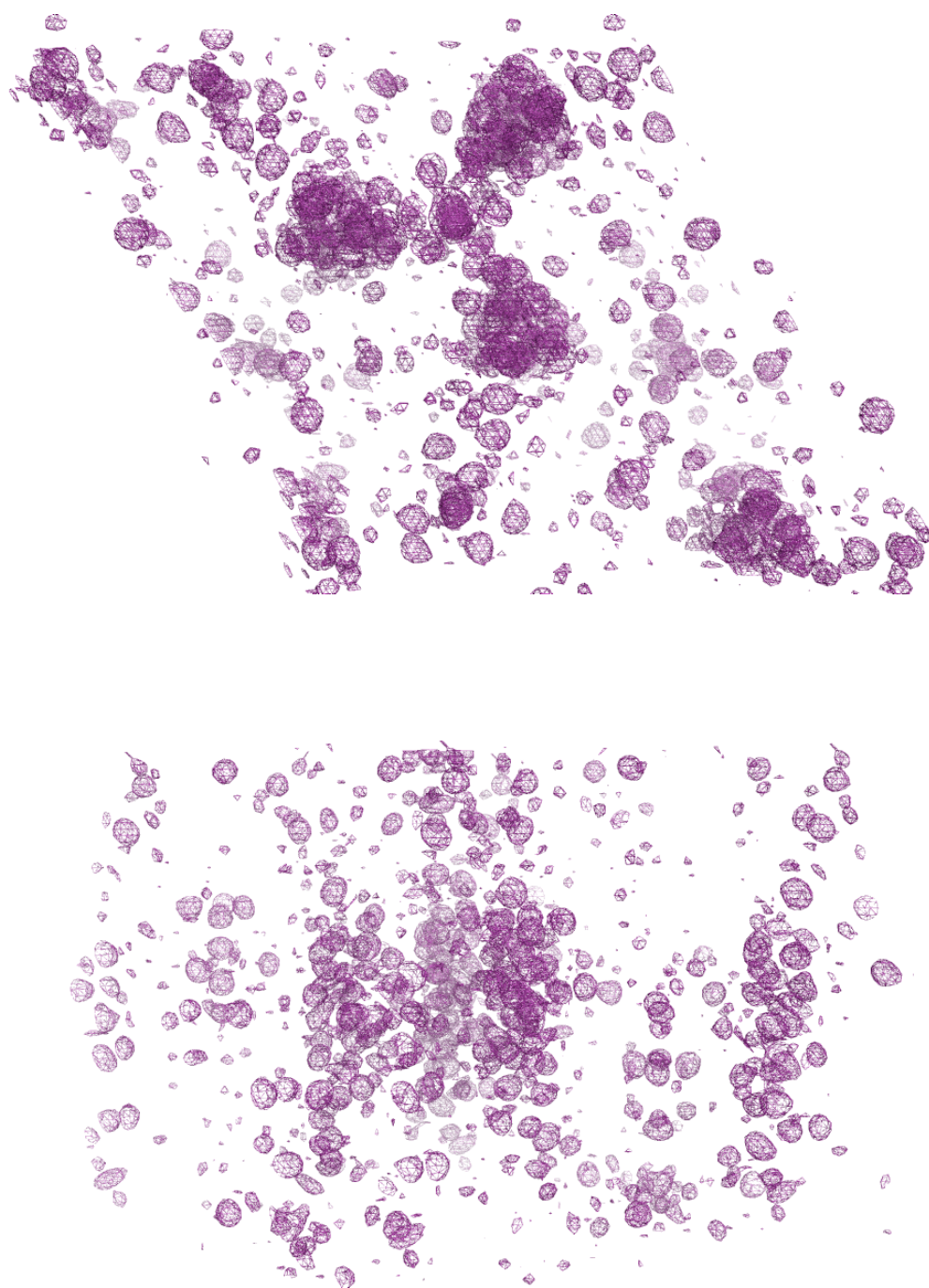


Figure 6.21: Electron density maps calculated *ab initio* using ACORN. The top view shows a view along the c-axis of the crystal, showing a clear region of electron density running roughly parallel to the axis. This corresponds to the predicted dimensions of an extended ZrCo molecule. The bottom view shows the same electron density but viewed from the side, perpendicular to the c-axis. The density in the central portion of the helix is relatively well-defined, but the density at the ends of the helix appears to splay outward, and becomes more diffuse.

6.7 Experimental phasing attempts

Alongside the computational approach to obtaining phases for the ZiCo1 structure, experimental phasing attempts were also carried out. Heavy atom compounds were selected on the basis of their reactivities at the given pH and other considerations. The compounds were dissolved in well solution (1M sodium citrate, 10mM sodium borate, pH8.5) at 10mM, and a small drop of the heavy atom soak solution was placed on a cover slip. A single ZiCo1 crystal was placed in each drop and left to soak for various lengths of time, from a few seconds to several hours, over a reservoir of well solution. The following compounds were tested at 10mM concentrations: K_2PtCl_4 , $\text{K}_2\text{Pt}(\text{NO}_2)_4$, $\text{K}_2\text{Pt}(\text{CN})_4$, $\text{KAu}(\text{CN})_4$, CdCl_2 and NaIrCl_6 .

In all conditions tested, the ZiCo1 crystals visibly deteriorated (see Figure 6.22). The observed cracking of crystals upon soaking with heavy atom compounds could be due to a number of factors: manipulation of the crystal; differences in pH or protein concentration between the mother liquor and the soak solution; or the chosen concentration of heavy atoms in the soak may simply have been too high.

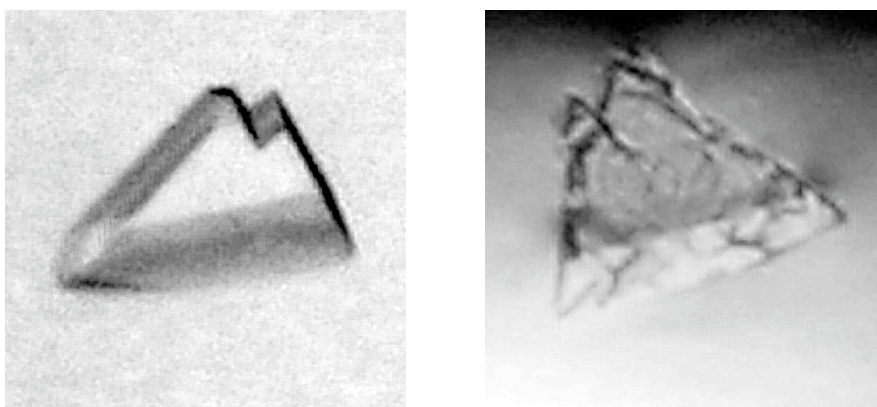


Figure 6.22: A crystal of the ZiCo1 before and after being soaked in a heavy atom solution. A single crystal of the peptide is shown (left). After soaking with well solution containing 10mM of various heavy atom compounds, the crystal visibly deteriorates and takes on a crazed appearance.

Mechanical stresses due to manipulation of the crystal are unlikely to be the cause of the observed cracking, since the crystals have proved to be stable when handled previously. Differences in pH between the natural mother liquor and the halide soak were ruled out, since both solutions had a pH of 7.9. The crystals were stable for several hours in the well solution, so the cracking was caused directly by the introduction of heavy atom compounds.

The experiments were repeated with 1mM heavy atom solutions, with similar results, although the crystals soaked in K_2PtCl_4 and $\text{K}_2\text{Pt}(\text{NO}_2)_4$ appeared to remain intact. These crystals were flash frozen without any additional cryoprotectant, and 180° of data were collected for each derivative. However, the data could not be processed due to high mosaic spread, suggesting that while the crystals may appear intact, they may have deteriorated on a microscopic level. This would seem to suggest that the introduction of heavy atoms is causing structural changes in the crystal that lead to its destruction. These may be due to the heavy atoms interfering with crystal contacts, or the induction of a conformational change in the ZiCo1 molecule on binding the heavy atom compounds.

Attempts at stabilising the crystals using higher concentrations of precipitant, adding a grain of heavy metal compound directly to the drop, and co-crystallising the peptide and heavy metal were unsuccessful. Since the computational attempts to solve the phase problem had yielded good results, experimental phasing using heavy atom compounds was abandoned as an approach to obtaining phase information for the ZiCo1 peptide.

6.8 Discussion

The goal of this project was to provide high-resolution structural information about the ZiCo1 and ZiCo95 peptides, in order to validate their design and gain an insight into their structural properties. In a wider context the aim of the ZiCo project is to gain an insight into the process of how to incorporate structural flexibility that bestows a novel function into the *de novo* design of a stably folded structure. Understanding the precise interactions that contribute to a successful protein design — which can only truly be achieved through high-resolution structures — will contribute to the development of robust and reliable methods of designing proteins with new and useful functions (Lassila, 2010).

Prior to the work described in this thesis, the ZiCo1 peptide had been crystallised using a peptide concentration of 2.25mM and 0.4M LiSO₄ as a precipitant, in the presence of 1mM zinc chloride at pH7.5. Since the crystallisation conditions contained significant zinc concentrations, it was proposed that the peptide had crystallised as a zinc finger. However, MR searches using a zinc finger model proved unsuccessful. Small helical fragments fared better, and it was initially assumed that these corresponded to the short alpha helix of the classical zinc finger structure. Alternatively, the crystal could contain the peptide in an unexpected conformation owing to the relatively low concentration of zinc in the crystallisation drop. In addition, the data quality left room for improvement (see Table 37), so new crystallisation conditions for ZiCo1 and an improved data set were implicit goals in this work.

	Data set 1			Data set 2		
Unit cell parameters	32.1	32.1	57.5	32.1	32.1	57.6
	90	90	120	90	90	120
Resolution	25.0–1.0			25.0–1.6		
Reflections	28951			71884		
Unique	25081			7661		
Completeness (overall–outer shell)	78–17.9 %			84.8–29.1 %		
Multiplicity (inner–outer shell)	2.8–1.3			11.1–3.5		
Rmerge (%)	5.5			4.8		
I/sd(I)	16.6			37.6		

Table 37: Data processing statistics for two ZiCo1 data sets. Data set 1 was collected by David Papapostolou and Nigel Brisset at the ESRF in Grenoble. Data set 2 was collected as described in section 6.3 ZiCo1 appears to have crystallised in the same unit cell and space group under both crystallisation conditions. While data set 1 extends to a maximum resolution of 1 Å, the second data set is more complete and has a lower R_{merge} .

Crystallisation of both peptides was attempted, but was unsuccessful for ZiCo95. Consequently, the solution structure of ZiCo95 was investigated using size exclusion chromatography. Analysis by SEC in a stabilising buffer showed that ZiCo95 exists as a heterogeneous population of oligomeric species, in both the presence and absence of zinc. The zinc-free peptide appears to form mostly trimers — as predicted by the design — with a small proportion forming tetramers. In the presence of zinc, ZiCo95 displays a much more heterogeneous population of oligomeric states. Most of the peptide is found as pentamers, followed by trimers, and finally a small proportion of monomers. When a reducing agent was included in the sample buffer, the peptide appeared to form a dimeric species that does not respond to zinc addition.

These results suggest that crystallising ZiCo95 in the zinc-bound form may be impossible, since a heterogeneous population of protein molecules is unfavourable for crystallisation. Furthermore, the behaviour in the presence of a reducing agent suggests that the peptide may not be folding as a zinc finger at all. A reducing agent causes any disulphide bonds between cysteine residues in the classical zinc finger motif to be

broken, leaving them free to co-ordinate zinc ions. However, since the peptide forms stable dimers in the presence of a reducing agent, it is possible that the peptides are associated via strong hydrophobic or electrostatic interactions. If these forces are stronger than the favourable dynamics exerted by zinc binding, this may explain why ZiCo95 does not respond to the addition of zinc under reducing conditions.

The ZiCo95 sequence contains a higher proportion of positively charged (K) and negatively charged residues (E) than the ZiCo1 sequence, and is more negatively charged at a given pH than ZiCo1. This fact lends support to the idea that electrostatic interactions between peptides may dominate over any decrease in entropy from adopting a coiled coil or zinc finger fold for ZiCo95. The notion that the peptide may not be correctly folded as either a trimeric coiled coil or a zinc finger is borne out by NMR studies of the peptide (Prof. Woolfson, personal communication). Therefore, the search for high-resolution structures of ZiCo95 was concluded, while the work with ZiCo1 continued.

It proved possible to crystallise the ZiCo1 peptide under a number of conditions. There was a noticeable difference in crystal morphology between conditions containing PEG as the primary precipitant and those containing salts, suggesting that electrostatic interactions may play a role in stabilising the ZiCo1 crystal structure. Large, single crystals of ZiCo1 were grown in the presence of equimolar zinc, using sodium citrate as the precipitant. These could be cryogenically frozen without additional cryoprotectant, and diffracted to 1.6 Å (see Table 37 for data processing statistics). Surprisingly, the crystals do not appear to contain zinc. This can be inferred by the absence of large, non-origin peaks in the native Patterson map, as would be expected to result from vectors between any electron-dense zinc atoms. Furthermore, no molecular replacement solutions could be found when a classical zinc finger was used as the search model, although molecular replacement with helical models gave more promising results.

Using a single 24-residue polyalanine helix in space group P3 gave the best results, although various different models were explored in all predicted space groups. These included a model featuring the key Phe10 residue (as well as numerous other key

residues) and trimeric coiled coils of varying lengths. The results of molecular replacement in P3 using the model described showed that ZiCo1 has apparently crystallised as a trimeric coiled coil. This was surprising, given the high concentrations of zinc required in order for the peptide to crystallise. This puzzling behaviour might be explained by the concentration-dependence exhibited by the peptide (see section 1.7.3).

The melting temperature of the trimer increases with the concentration of peptide in solution, suggesting that as the peptide concentration increases, the position of equilibrium is shifted in favour of the trimeric form over a random coil in the absence of zinc. The ZiCo1 crystallisation conditions initially show precipitated protein, with crystals being observed a few days later. This suggests that the combination of zinc and precipitant causes enough of the peptide (and zinc) to be removed from the solution to allow trimeric nuclei to form. As peptide is incorporated by the growing crystals, the precipitated protein is consumed. However, if this unique process is what drives the crystallisation of the ZiCo1 peptide, it may be impossible to crystallise the zinc-bound form. Nonetheless, the results of the molecular replacement corroborate the previous biophysical studies carried out by others that suggest that the ZiCo1 peptide folds as a trimeric coiled coil.

In the MR solution in space group P3, the helices making up the coiled coil are related by crystallographic symmetry, and there are two helices in the asymmetric unit. A similar solution was found using a trimeric coiled coil model by searching in space group P1, which has no crystallographic symmetry elements. This placed six individual helices in the unit cell, as two trimeric coiled coils related by a non-crystallographic symmetry axis running parallel to the long axis of the cell. In addition, all the ZiCo1 crystals tested exhibit signs of being almost perfectly twinned. This is demonstrated by the results of twinning tests, leading to considerable ambiguity about the true space group of the crystal.

The predicted space group is P312, which relates molecules as shown in Appendix II, section ii. This means a single helix occupies the asymmetric unit, and six copies are related by crystallographic symmetry in the unit cell. However, MR searches

in P312 failed to produce any solutions because of steric clashes between placed helices. A further insight into the presence of crystallographic and non-crystallographic symmetry elements in the ZiCo1 data came from the self rotation function. Calculating the self rotation function in the various possible space groups showed the presence of numerous symmetry operators (see Appendix II, section iii). Given that an MR solution is found in P1, where two crystallographic trimers are related by a non-crystallographic symmetry axis, it is possible that the apparent P312 symmetry could be explained by the presence of NCS or twinning that mimics this space group. Therefore the true symmetry of the crystal is likely to be P3, or possibly even P1 with a high degree of non-crystallographic symmetry, and the additional possible complication of twinning.

The output of the molecular replacement procedure in space group P3 was used as a starting point for automated model building, producing a model with 99% completeness and an R-factor of 0.35. Further refinement gave a final R-factor of 0.23, which is consistent with a well-refined structure. The slight divergence between the R-factor and R_{free} during the final rounds of refinement may be explained by the incompleteness of the model. Some residues could not be built using the automated procedure due to steric clashes, and these may be disordered residues at the ends of the helix. While the results of molecular replacement are biased towards the initial model used in the search, the results of *ab initio* phasing also revealed a trimeric, helical structure. This supports the conclusion that ZiCo1 has indeed crystallised in its trimeric form, and that the refined model may be largely correct.

The packing of the final model in the unit cell is shown in Figure 6.23. ZiCo1 trimers are formed by the combination of symmetry operators and the crystalline lattice, and appear to be stacked in antiparallel layers. An interesting feature that can be observed in the final structure is that His6 appears to be co-ordinating an atom (see Figure 6.24). There is a strong blob of density that persists when the maps are contoured at high sigma, which suggests that an electron is highly likely to be located in this area. However, this may not be significant, and the solution may not be entirely correct, given the possible ambiguity of the space group.

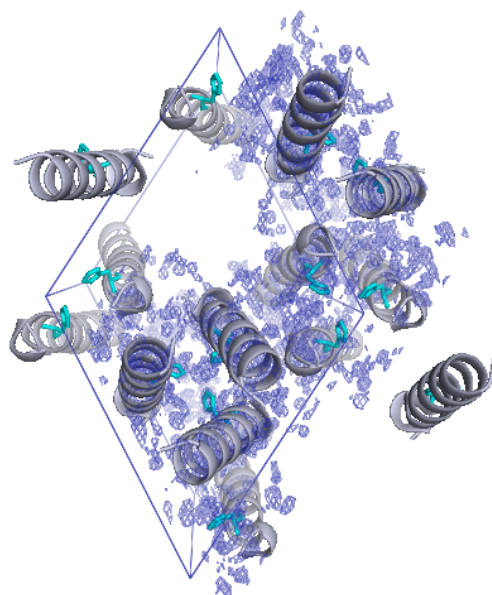


Figure 6.23: Packing of the refined ZiCo1 solution in space group P3. The unit cell is indicated and the view is along the long cell edge, which corresponds to the coiled coil axis. Helices are represented as cartoon ribbons, with His6 highlighted.

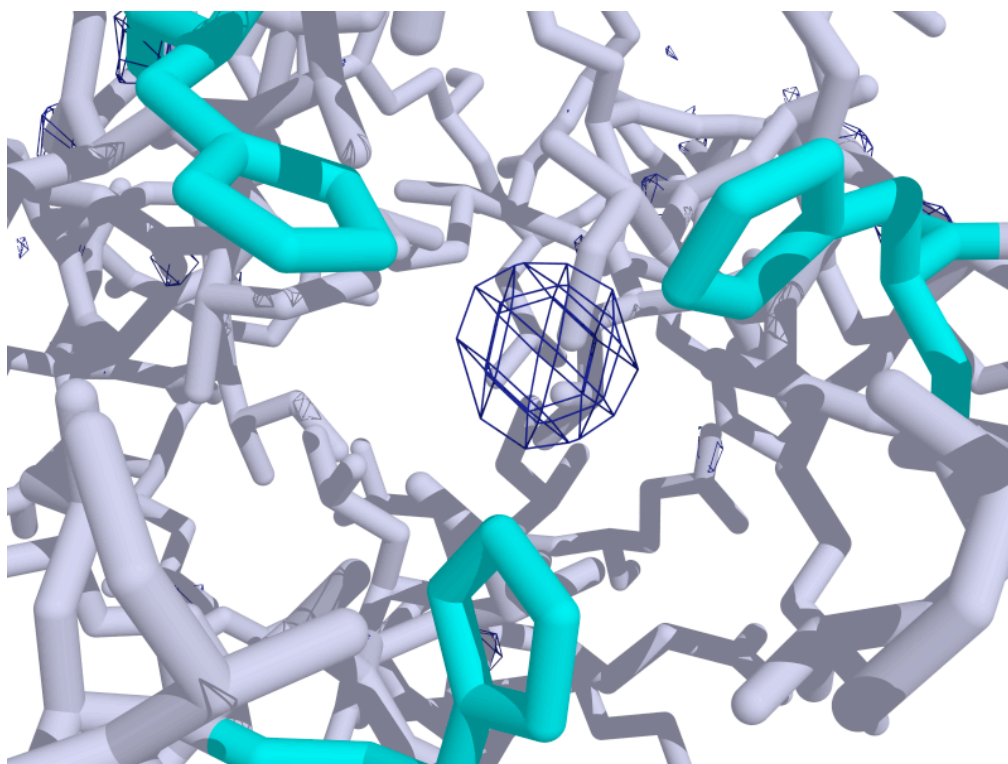


Figure 6.24: Electron density near His6 in the refined ZiCo1 structure. The map is contoured at 5 σ , and shows a persistent blob of density between three symmetry-related histidines at position 6 in the ZiCo1 sequence.

There is scope for the project to be developed further, with possible work including further manual model building and refinement to improve the structure (for example by refining in space group P1), or treating the potentially twinned data more effectively by carrying out twin refinement using SHELX (Sheldrick, 2008). However, the aims set out for this project have been adequately fulfilled for the coiled coil form of ZiCo1. The ZiCo1 peptide has been crystallised as a trimeric coiled coil, which is likely to reflect the structure of the peptide in solution. In terms of validating the design, the ZiCo1 sequence was designed in such a way that it would fold as a coiled-coil type structure, and therefore this aspect of the design process has been successful. Considering that the zinc finger form — which is more interesting from a protein design perspective — could not be crystallised, this project has reached a state of completion which is in accordance with the aims of the project and the available time and resources.

Chapter 7

Conclusions

7.1 Conclusions

7.1.1 Complement component C1

A novel purification scheme to isolate native complement component C1 from human plasma has been developed, using a combination of PEG precipitation, size exclusion and anion exchange chromatography. All of the C1 subcomponents were shown to be enriched in the purified sample by SDS-PAGE and mass spectrometry. Dynamic light scattering studies showed a single population of particles around 40nm in diameter, which corresponds to the estimated size of the C1 complex. Therefore C1 has been purified by a new procedure in a form suitable for crystallisation. In addition, investigations into the effects of relevant metal ions on the structure of the purified complex under physiologically relevant conditions were pursued. Finally, the crystallisability of complement component C1 was explored with the hope of achieving usable crystals. The results suggest that C1 exhibits a difference in solubility when salts are used as precipitants compared with polyethylene glycols, which may be useful information in the search for crystallisation conditions. Several promising conditions were identified from preliminary screens, and so the objectives set out for this project can be deemed to have been met successfully.

7.1.2 Serum amyloid P component

With regards to SAP, the aim of my work was to perform preliminary investigations into the metal- and ligand-binding properties of this molecule in order to lay the foundations for further work. SAP was purified from human plasma using established methods and the quality of the purified product was analysed for its suitability for use in structural studies. SAP was found to be pure and homogeneous by DLS and could be crystallised under known conditions. The metal-binding properties of SAP were examined using the ability of the ligand PE to disrupt the calcium-dependent aggregation of SAP. This interaction may have functional implications for the role of SAP in immunity, and the simple assay described here may offer a quick and simple means of investigating the metal- and ligand-binding properties of SAP. Furthermore, a possible common functional mechanism between SAP and C1 may have been identified, since both proteins exhibit copper-dependent aggregation at the low pH characteristic of sites of infection and inflammation. Further work on this topic might include studying the interaction of SAP with copper, zinc and physiological ligands such as PE at acidic and neutral pH, and co-crystallisation of SAP with various divalent metal ions in order to more fully determine the role of metals in the structure and function of serum amyloid P component.

7.1.3 ZiCo

The aims set out for this project have been adequately fulfilled for the coiled coil form of ZiCo1. The peptide has been crystallised as a trimeric coiled coil, which is likely to reflect the structure of the peptide in solution. While the apparent space group is P312, it is likely that the true space group may be a lower-symmetry space group with NCS and/or twinning. However, both MR and *ab initio* phasing suggested a coiled-coil type structure within the crystal, which could be traced by an automated model building procedure. This helical structure was refined to a final R-factor of 0.23, which is consistent with a well-refined structure. In terms of validating the design, it the ZiCo1 sequence was designed in such a way that it would fold as a coiled-coil type structure, and therefore this aspect of the design process has been successful. While neither the zinc-finger form of ZiCo1, nor the ZiCo95 peptide could be crystallised, the pursuit of high-resolution structures of ZiCo has reached a state of completion in accordance with the aims set out for the project.

7.2 Further work

The C1 crystals obtained were too small to be tested for diffraction, so there is considerable scope for further work on the crystallisation of complement component C1. In addition, the purified C1 sample may yet contain other complement-related proteins and high-abundance serum proteins, so further purification or alternative purification methods could also be explored. Mono Q-purified C1 may also be a more suitable sample than SEC-purified C1 for electron microscopy. Finally, the interaction of purified C1 with divalent metal cations — and also C1 inhibitor — merit further investigation. However, the aims set out for this project have been adequately fulfilled.

Further work on SAP might include studying the interaction with copper, zinc, and physiological ligands such as PE at acidic and neutral pH. This could be accomplished either using DLS and the assay method described in this work, or by other techniques such as CD and FTIR. The co-crystallisation of SAP with various divalent metal ions and small-molecule ligands may also shed some light on the role of divalent metal cations in modulating the structural and functional properties of this interesting immune protein. The aims of this project have been fulfilled, in that further work has already been carried out by others on the basis of the work outlined herein.

There is also scope for the ZiCo project to be developed further, with possible avenues including further manual model building and refinement to improve the structure (for example by attempting refinement in space group P1), or treating the potentially twinned data more effectively by carrying out twin refinement using alternative programs. However, the zinc finger form is more interesting in terms of validating the protein design, and this could not be crystallised. Furthermore, subsequent NMR studies on the peptide also proved challenging (Dek Woolfson, personal communication), and so this project has been deemed completed.

7.3 Concluding remarks

This final section will enclose the present thesis in its wider context. The importance of high-resolution protein structures has been introduced and the role of metals and metal-binding proteins in the cause and treatment of disease have been explored — with specific reference to complement component C1, serum amyloid P component and the designed peptide ZiCo. Salient structural features of these proteins and our current understanding of their respective functions have been presented and critically discussed. The background to my work has also been explored and I believe I have demonstrated how this thesis complements previous work on the three proteins under study and opens up new avenues of research. Finally, the aims set out for these investigations into the structures of the serum proteins complement component C1, serum amyloid P component, and the designed switch peptide ZiCo have been acceptably met.

Bibliography

Abel, K. et al. (1996) An Alpha to Beta Conformational Switch in EF-Tu. *Structure*, **4**, 1153-1159.

Ambroggio, X.I. & Kuhlman, B. (2006) Computational design of a single amino acid sequence that can switch between two distinct protein folds. *J Am Chem Soc*, **128**, 1154-1161.

Anfinsen, C.B. (1973) Principles That Govern the Folding of Protein Chains. *Science*, **181**, 223-230.

Anfinsen, C.B. & Haber, E. (1961) Studies on the Reduction and Re-formation of Protein Disulfide Bonds. *J Biol Chem*, **236**, 1361-1363.

Arlaud, G.J. et al. (1980) A study on the structure and interactions of the C1 sub-components C1r and C1s in the fluid phase. *Biochim Biophys Acta*, **616**, 105-115.

Arlaud, G.J. et al. (2002) Structural biology of the C1 complex of complement unveils the mechanisms of its activation and proteolytic activity. *Mol Immunol*, **39**, 383-394.

Arlaud, G.J. et al. (1980) Purified proenzyme C1r. Some characteristics of its activation and subsequent proteolytic cleavage. *Biochim Biophys Acta*, **616**, 116-129.

Asakura, T., Adachi, K. & Schwartz, E. (1978) Stabilizing effect of various organic solvents on protein. *J Biol Chem*, **253**, 6423-6425.

Ashton, A.W. et al. (1997) Pentameric and decameric structures in solution of serum amyloid P component by X-ray and neutron scattering and molecular modelling analyses. *J Mol Biol*, **272**, 408-422.

Basharov, M.A. (2000) The Posttranslational Concept of Protein Folding: How valid is it? *Biochemistry (Mosc)*, **65**, 1184-1191.

Basharov, M.A. (2003) Protein Folding. *J Cell Mol Med*, **7**, 223-237.

Beinrohr, L. et al. (2008) C1, MBL-MASPs and C1-inhibitor: novel approaches for targeting complement-mediated inflammation. *Trends Mol Med*, **14**, 511-521.

Berg, J.M. & Godwin, H.A. (1997) Lessons from zinc-binding peptides. *Annu Rev Biophys Biomol Struct*, **26**, 357-371.

Bickerstaff, M.C. et al. (1999) Serum amyloid P component controls chromatin degradation and prevents antinuclear autoimmunity. *Nat Med*, **5**, 694-697.

Bijl, M. et al. (2003) Serum amyloid P component binds to late apoptotic cells and mediates their uptake by monocyte-derived macrophages. *Arthritis Rheum*, **48**, 248-254.

- Bolen, D.W. (2004) Effects of naturally occurring osmolytes on protein stability and solubility: issues important in protein crystallization. *Methods*, **34**, 312-322.
- Bollag, D.M., Rozycki, M.D. & Edelstein, S.J. (1996) *Protein methods*. Wiley-Liss, New York.
- Bottazzi, B. et al. (2006) Pentraxins as a key component of innate immunity. *Curr Opin Immunol*, **18**, 10-15.
- Brunger, A.T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472-475.
- Budayova-Spano, M. et al. (2002) Monomeric structures of the zymogen and active catalytic domain of complement protease c1r: further insights into the c1 activation mechanism. *Structure*, **10**, 1509-1519.
- Candiano, G. et al. (2004) Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis*, **25**, 1327-1333.
- Carrell, R.W. & Gooptu, B. (1998) Conformational Changes and Disease – Serpins, Prions and Alzheimer's. *Curr Opin Struct Biol*, **8**, 799-809.
- CCP4 Wiki. Accessed 26 February 2011. <http://ccp4wiki.org/~ccp4wiki/wiki/index.php?title=Twining#The_warning_signs_for_twinning>
- Cerasoli, E., Sharpe, B.K. & Woolfson, D.N. (2005) ZiCo: a peptide designed to switch folded state upon binding zinc. *J Am Chem Soc*, **127**, 15008-15009.
- Cerpa, R., Cohen, F.E. & Kuntz, I.D. (1996) Conformational Switching in Designed Peptides: the helix/sheet transition. *Fold Des*, **1**, 91-101.
- Chayen, N.E. (2004) Turning protein crystallisation from an art into a science. *Curr Opin Struct Biol*, **14**, 577-583.
- Chayen, N.E., Saridakis, E. & Sear, R.P. (2006) Experiment and theory for heterogeneous nucleation of protein crystals in a porous medium. *Proc Natl Acad Sci U S A*, **103**, 597-601.
- Chernov, A.A. (2003) Protein crystals and their growth. *J Struct Biol*, **142**, 3-21.
- Chevallet, M., Luche, S. & Rabilloud, T. (2006) Silver staining of proteins in polyacrylamide gels. *Nat Protoc*, **1**, 1852-1858.
- Ciani, B. et al. (2002) A Designed System for Assessing How Sequence Affects Alpha to Beta Conformational Transitions in Proteins. *J Biol Chem*, **277**, 10150-10155.
- Clemenza, L. et al. (2003) Research on complement: old issues revisited and a novel sphere of influence. *Trends Immunol*, **24**, 292-296.

- Cohn, E.J., Hughes, W.L.J. & Weare, J.H. (1947) Preparation and properties of serum and plasma proteins; crystallization of serum albumins from ethanol water mixtures. *J Am Chem Soc*, **69**, 1753-1761.
- Collaborative Computational Project, N. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*, **50**, 760-763.
- Collins, K.D. (2004) Ions from the Hofmeister series and osmolytes: effects on proteins in solution and in the crystallization process. *Methods*, **34**, 300-311.
- Colten, H.R., Borsos, T. & Rapp, H.J. (1968) Reversible loss of activity of the first component of complement (C'1) as a function of ionic strength. *J Immunol*, **100**, 799-807.
- Cooper, N.R. (1985) The classical complement pathway: activation and regulation of the first complement component. *Adv Immunol*, **37**, 151-216.
- Cudney, B. (1999) Protein Crystallization and Dumb Luck. *The Rigaku Journal*, **16**, 1-7.
- D'Andrea, M.R. (2005) Evidence that immunoglobulin-positive neurons in Alzheimer's disease are dying via the classical antibody-dependent complement pathway. *Am J Alzheimers Dis Other Demen*, **20**, 144-150.
- Damaschun, G. et al. (1999) Proteins Can Adopt Totally Different Folded Conformations. *J Mol Biol*, **291**, 715-725.
- Danielsen, B. et al. (1997) Calcium-dependent and -independent binding of the pentraxin serum amyloid P component to glycosaminoglycans and amyloid proteins: enhanced binding at slightly acid pH. *Biochim Biophys Acta*, **1339**, 73-78.
- Davis, A.E.R., Mejia, P. & Lu, F. (2008) Biological activities of C1 inhibitor. *Mol Immunol*, **45**, 4057-4063.
- De Beer, F.C. & Pepys, M.B. (1982) Isolation of human C-reactive protein and serum amyloid P component. *J Immunol Methods*, **50**, 17-31.
- De Zwaan, C., van Dieijen-Visser, M.P. & Hermens, W.T. (2003) Prevention of cardiac cell injury during acute myocardial infarction: possible role for complement inhibition. *Am J Cardiovasc Drugs*, **3**, 245-251.
- Dias, Joao M. 2002. Re: [ccp4bb]:SUMMARY- unusually high solvent content. Accessed 12 August 2009. <<http://www.ysbl.york.ac.uk/ccp4bb/2002/msg00205.html>>
- Diem, K. & Lentner, C. (1975) *Scientific Tables*. Geigy Pharmaceuticals, Ardsley, NY.
- Dill, K.A. et al. (2008) The protein folding problem. *Annu Rev Biophys*, **37**, 289-316.

- Dong, A., Caughey, W.S. & Du Clos, T.W. (1994) Effects of calcium, magnesium, and phosphorylcholine on secondary structures of human C-reactive protein and serum amyloid P component observed by infrared spectroscopy. *J Biol Chem*, **269**, 6424-6430.
- Doubl  , S. (2007) *Macromolecular crystallography protocols*. Humana Press, Totowa, N.J.
- Drenth, J. (2002) *Principles of Protein X-ray Crystallography (Springer Advanced Texts in Chemistry)*. Springer,
- Duncan, R.C., Wijeyewickrema, L.C. & Pike, R.N. (2008) The initiating proteases of the complement system: controlling the cleavage. *Biochimie*, **90**, 387-395.
- Durbin, S.D. & Feher, G. (1996) Protein crystallization. *Annu Rev Phys Chem*, **47**, 171-204.
- Eaton, W.A. et al. (2000) Fast Kinetics and Mechanisms in Protein Folding. *Annu Rev Biophys Biomol Struct*, **29**, 327-359.
- Ellis, R.J. (2006) Molecular Chaperones: Assisting Assembly in Addition to Folding. *Trends Biochem Sci*, **31**, 395-401.
- Emsley, J. et al. (1994) Structure of pentameric human serum amyloid P component. *Nature*, **367**, 338-345.
- Emsley, P. & Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, **60**, 2126-2132.
- Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr*, **62**, 72-82.
- Evans, P. & McCoy, A. (2008) An introduction to molecular replacement. *Acta Crystallogr D Biol Crystallogr*, **64**, 1-10.
- Familian, A. et al. (2001) Chromatin-independent binding of serum amyloid P component to apoptotic cells. *J Immunol*, **167**, 647-654.
- Fisher, T.E. et al. (1999) The Study of Protein Mechanics With the Atomic Force Microscope. *Trends Biochem Sci*, **24**, 379-384.
- Foadi, J. et al. (2000) A flexible and efficient procedure for the solution and phase refinement of protein structures. *Acta Crystallographica Section D*, **56**, 1137-1147.
- French, S. & Wilson, K. (1978) On the treatment of negative intensity observations. *Acta Crystallographica Section A*, **34**, 517-525.
- Garcia de Frutos, P. & Dahlback, B. (1994) Interaction between serum amyloid P component and C4b-binding protein associated with inhibition of factor I-mediated C4b degradation. *J Immunol*, **152**, 2430-2437.

- Gasque, P. (2004) Complement: a unique innate immune sensor for danger signals. *Mol Immunol*, **41**, 1089-1098.
- Gewurz, H., Zhang, X.H. & Lint, T.F. (1995) Structure and function of the pentraxins. *Curr Opin Immunol*, **7**, 54-64.
- Gilliland, G.L. et al. (1994) Biological Macromolecule Crystallization Database, Version 3.0: new features, data and the NASA archive for protein crystal growth data. *Acta Crystallogr D Biol Crystallogr*, **50**, 408-413.
- Gingras, A.C. et al. (2007) Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol*, **8**, 645-654.
- Gregory, L.A. et al. (2003) X-ray structure of the Ca²⁺-binding interaction domain of C1s. Insights into the assembly of the C1 complex of complement. *J Biol Chem*, **278**, 32157-32164.
- Hanlon, D.P., Watt, D.S. & Westhead, E.W. (1966) The interaction of divalent metal ions with tris buffer in dilute solution. *Anal Biochem*, **16**, 225-233.
- Hartl, F.U. & Hayer-Hartl, M. (2009) Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol*, **16**, 574-581.
- Heck, A.J. (2008) Native mass spectrometry: a bridge between interactomics and structural biology. *Nat Methods*, **5**, 927-933.
- Heegaard, N.H., He, X. & Blomberg, L.G. (2006) Binding of Ca²⁺, Mg²⁺, and heparin by human serum amyloid P component in affinity capillary electrophoresis. *Electrophoresis*, **27**, 2609-2615.
- Hicks, P.S. et al. (1992) Serum amyloid P component binds to histones and activates the classical complement pathway. *J Immunol*, **149**, 3689-3694.
- Hori, Y. et al. (2000) The engineering, structure, and DNA binding properties of a novel His4-type zinc finger peptide. *Nucleic Acids Symp Ser*, 295-296.
- Huntington, J.A., Read, R.J. & Carrell, R.W. (2000) Structure of a serpin-protease complex shows inhibition by deformation. *Nature*, **407** (6806), 923-926.
- Hutchinson, W.L., Hohenester, E. & Pepys, M.B. (2000) Human serum amyloid P component is a single uncomplexed pentamer in whole serum. *Mol Med*, **6**, 482-493.
- Kastenholz, B. (2007) New hope for the diagnosis and therapy of Alzheimer's disease. *Protein Pept Lett*, **14**, 389-393.
- Keller, A. et al. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem*, **74**, 5383-5392.

- Kerr, F.K. et al. (2008) Elucidation of the substrate specificity of the MASP-2 protease of the lectin complement pathway and identification of the enzyme as a major physiological target of the serpin, C1-inhibitor. *Mol Immunol*, **45**, 670-677.
- Kishore, U. & Reid, K.B. (2000) C1q: structure, function, and receptors. *Immunopharmacology*, **49**, 159-170.
- Koles, K. et al. (2004) N- and O-glycans of recombinant human C1 inhibitor expressed in the milk of transgenic rabbits. *Glycobiology*, **14**, 51-64.
- Lassila, J.K. (2010) Conformational diversity and computational enzyme design. *Curr Opin Chem Biol*, **14**, 676-682.
- Leone, N. et al. (2006) Zinc, copper, and magnesium and risks for all-cause, cancer, and cardiovascular mortality. *Epidemiology*, **17**, 308-314.
- Lepow, I.H. et al. (1963) Chromatographic resolution of the first component of human complement into three activities. *J Exp Med*, **117**, 983-1008.
- Lepow, I.H., Ratnoff, O.D. & Levy, L.R. (1958) Studies on the activation of a proesterase associated with partially purified first component of human complement. *J Exp Med*, **107**, 451-474.
- Leslie, A.G. (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, **26**.
- Levinthal, C. (1968) Are There Pathways for Protein Folding? *J Chim Phys*, **65**, 44-45.
- Li, X.A. et al. (1995) Characterization of serum amyloid P component from human aortic atherosclerotic lesions. *Arterioscler Thromb Vasc Biol*, **15**, 252-257.
- Lippi, M. et al. (2008) MetalDetector: a web server for predicting metal-binding sites and disulfide bridges in proteins from sequence. *Bioinformatics*, **24**, 2094-2095.
- Lippow, S.M. & Tidor, B. (2007) Progress in computational protein design. *Curr Opin Biotechnol*, **18**, 305-311.
- Long, D. & Yang, D. (2009) Buffer interference with protein dynamics: a case study on human liver fatty acid binding protein. *Biophys J*, **96**, 1482-1488.
- Mac Sweeney, A. & D'Arcy, A. (2003) A simple and rapid method for mounting protein crystals at room temperature. *Journal of Applied Crystallography*, **36**, 165-166.
- Malek, F. et al. (2006) Serum copper as a marker of inflammation in prediction of short term outcome in high risk patients with chronic heart failure. *Int J Cardiol*, **113**(2), e51-3.

- Mantovani, A. et al. (2008) Pentraxins in innate immunity: from C-reactive protein to the long pentraxin PTX3. *J Clin Immunol*, **28**, 1-13.
- Manuel Garcia-Ruiz, J. (2003) Nucleation of protein crystals. *J Struct Biol*, **142**, 22-31.
- McCoy, A.J. (2004) Liking likelihood. *Acta Crystallogr D Biol Crystallogr*, **60**, 2169-2183.
- McCoy, A.J. et al. (2007) Phaser crystallographic software. *Journal of Applied Crystallography*, **40**, 658-674.
- McPherson, A. (2004) Introduction to protein crystallization. *Methods*, **34**, 254-265.
- McRee, D.E. (1999) *Practical protein crystallography*. Academic Press, San Diego, Calif.; London.
- Medicus, R.G. & Chapuis, R.M. (1980) The first component of complement. I. Purification and properties of native C1. *J Immunol*, **125**, 390-395.
- Merril, C.R. et al. (1988) Coloration of silver-stained protein bands in polyacrylamide gels is caused by light scattering from silver grains of characteristic sizes. *Proc Natl Acad Sci U S A*, **85**, 453-457.
- Mezei, M. (1998) Chameleon Sequences in the PDB. *Protein Eng*, **11**, 411-414.
- Minor, D.J. & Kim, P. (1996) Context-dependent Secondary Structure Formation of a Designed Protein Sequence. *Nature*, **380**, 730-734.
- Mold, C., Baca, R. & Du Clos, T.W. (2002) Serum amyloid P component and C-reactive protein opsonize apoptotic cells for phagocytosis through Fcgamma receptors. *J Autoimmun*, **19**, 147-154.
- Molina, H. (2004) Complement and immunity. *Rheum Dis Clin North Am*, **30**, 1-18, v.
- Mollnes, T.E. & Kirschfink, M. (2006) Strategies of therapeutic complement inhibition. *Mol Immunol*, **43**, 107-121.
- Morgan, B.P. et al. (2005) Complement: central to innate immunity and bridging to adaptive responses. *Immunol Lett*, **97**, 171-179.
- Nave, C. & Garman, E.F. (2005) Towards an understanding of radiation damage in cryocooled macromolecular crystals. *J Synchrotron Radiat*, **12**, 257-260.
- Neoh, S.H., Gordon, T.P. & Roberts-Thomson, P.J. (1984) A simple one-step procedure for preparation of C1-deficient human serum. *J Immunol Methods*, **69**, 277-280.
- Nesvizhskii, A.I. et al. (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem*, **75**, 4646-4658.

- Nicholson-Weller, A. & Klickstein, L.B. (1999) C1q-binding proteins and C1q receptors. *Curr Opin Immunol*, **11**, 42-46.
- Niculescu, F. & Rus, H. (1999) Complement activation and atherosclerosis. *Mol Immunol*, **36**, 949-955.
- Novitskaya, V. et al. (2006) Amyloid fibrils of mammalian prion protein are highly toxic to cultured cells and primary neurons. *J Biol Chem*, **281**, 13828-13836.
- OED Online. 2008. "protein, n.". *Oxford University Press*. Accessed 16/09/2008. <<http://dictionary.oed.com.ezproxy.sussex.ac.uk/cgi/entry/50190693>>
- Ogden, C.A. & Elkon, K.B. (2006) Role of complement and other innate immune mechanisms in the removal of apoptotic cells. *Curr Dir Autoimmun*, **9**, 120-142.
- Ohshima, S. et al. (2003) Galectin 3 and its binding protein in rheumatoid arthritis. *Arthritis Rheum*, **48**, 2788-2795.
- Oksjoki, R. et al. (2007) Function and regulation of the complement system in cardiovascular diseases. *Front Biosci*, **12**, 4696-4708.
- Owen, C.A.J. (1975) Uptake of ^{67}Cu by ceruloplasmin in vitro. *Proc Soc Exp Biol Med*, **149**, 681-682.
- Paidassi, H. et al. (2008) C1q binds phosphatidylserine and likely acts as a multiligand-bridging molecule in apoptotic cell recognition. *J Immunol*, **180**, 2329-2338.
- Pandya, M.J. et al. (2004) Sequence and Structural Duality: Designing Peptides to Adopt Two Stable Conformations. *J Am Chem Soc*, **126**, 17016-17024.
- Pearce, D.A., Walkup, G.K. & Imperiali, B. (1998) Peptidyl chemosensors incorporating a FRET mechanism for detection of Ni(II). *Bioorg Med Chem Lett*, **8**, 1963-1968.
- Pepys, M.B. et al. (1997) Amyloid P component. A critical review. *Amyloid: International Journal of Experimental & Clinical Investigation*, **4**, 274-295.
- Pepys, M.B. (1974) Role of complement in induction of antibody production in vivo. Effect of cobra factor and other C3-reactive agents on thymus-dependent and thymus-independent antibody responses. *J Exp Med*, **140**, 126-145.
- Pepys, M.B. et al. (1979) Binding of serum amyloid P-component (SAP) by amyloid fibrils. *Clin Exp Immunol*, **38**, 284-293.
- Pepys, M.B. & Butler, P.J. (1987) Serum amyloid P component is the major calcium-dependent specific DNA binding protein of the serum. *Biochem Biophys Res Commun*, **148**, 308-313.
- Pepys, M.B. et al. (2002) Targeted pharmacological depletion of serum amyloid P component for treatment of human amyloidosis. *Nature*, **417**, 254-259.

Perkins, S.J. (1985) Molecular modelling of human complement subcomponent C1q and its complex with C1r2C1s2 derived from neutron-scattering curves and hydrodynamic properties. *Biochem J*, **228**, 13-26.

Perrakis, A. et al. (2001) ARP/wARP and molecular replacement. *Acta Crystallogr D Biol Crystallogr*, **57**, 1445-1450.

Potempa, L.A., Kubak, B.M. & Gewurz, H. (1985) Effect of divalent metal ions and pH upon the binding reactivity of human serum amyloid P component, a C-reactive protein homologue, for zymosan. Preferential reactivity in the presence of copper and acidic pH. *J Biol Chem*, **260**, 12142-12147.

Prodeus, A.P. et al. (1998) A critical role for complement in maintenance of self-tolerance. *Immunity*, **9**, 721-731.

Radaev, S., Li, S. & Sun, P.D. (2006) A survey of protein-protein complex crystallizations. *Acta Crystallogr D Biol Crystallogr*, **62**, 605-612.

Radford, S.E. (2000) Protein Folding: Progress Made and Promises Ahead. *Trends Biochem Sci*, **25**, 611-618.

Regal, J.F. (1997) Role of the complement system in pulmonary disorders. *Immunopharmacology*, **38**, 17-25.

Reid, K.B. (1979) Complete amino acid sequences of the three collagen-like regions present in subcomponent C1q of the first component of human complement. *Biochem J*, **179**, 367-371.

Reid, K.B. & Porter, R.R. (1976) Subunit composition and structure of subcomponent C1q of the first component of human complement. *Biochem J*, **155**, 19-23.

Rosenberg, B. et al. (1967) The inhibition of growth or cell division in *Escherichia coli* by different ionic species of platinum(IV) complexes. *J Biol Chem*, **242**, 1347-1352.

Ross, R. (1999) Atherosclerosis is an inflammatory disease. *Am Heart J*, **138**, S419-20.

Rother, K., Till, G.O. & Hänsch, G.M. (1998) The complement system.

Sacks, S.H., Chowdhury, P. & Zhou, W. (2003) Role of the complement system in rejection. *Current Opinion in Immunology*, **15**, 487-492.

Sakamoto, M., Fujisawa, Y. & Nishioka, K. (1998) Physiologic role of the complement system in host defense, disease, and malnutrition. *Nutrition*, **14**, 391-398.

Schenk, H. (1984) *An introduction to direct methods : the most important phase relationships and their application in solving the phase problem*. Published for the International Union of Crystallography by University College Cardiff Press, Cardiff, Wales.

- Schifferli, J.A. & Steiger, G. (1985) A simple two-step procedure for the preparation of the first component of human complement (C1) in its native form. *J Immunol Methods*, **76**, 283-288.
- Schumaker, V.N. et al. (1981) Semi-flexible joint in the C1q subunit of the first component of human complement. *J Mol Biol*, **148**, 191-197.
- Sharon, M. & Robinson, C.V. (2007) The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annu Rev Biochem*, **76**, 167-193.
- Sheldrick, G.M. (2008) A short history of SHELX. *Acta Crystallogr A*, **64**, 112-122.
- Shi, W. & Chance, M.R. (2008) Metallomics and metalloproteomics. *Cell Mol Life Sci*, **65**, 3040-3048.
- Shintake, T. (2008) Possibility of single biomolecule imaging with coherent amplification of weak scattering x-ray photons. *Phys Rev E Stat Nonlin Soft Matter Phys*, **78**, 041906.
- Shults, M.D., Pearce, D.A. & Imperiali, B. (2003) Modular and tunable chemosensor scaffold for divalent zinc. *J Am Chem Soc*, **125**, 10591-10597.
- Signarvic, R.S. & DeGrado, W.F. (2003) De Novo Design of a Molecular Switch: Phosphorylation-dependent Association of Designed Peptides. *J Mol Biol*, **334**, 1-12.
- Sjoholm, A.G. et al. (2006) Complement deficiency and disease: an update. *Mol Immunol*, **43**, 78-85.
- Skarnes, R.C. & Watson, D.W. (1957) Antimicrobial factors of normal tissues and fluids. *Bacteriol Rev*, **21**, 273-294.
- Smith, L.J. et al. (1996) The Concept of a Random Coil. Residual Structure in Peptides and Denatured Proteins. *Fold Des*, **1**, R95-106.
- Sohl, J.L., Jaswal, S.S. & Agard, D.A. (1998) Unfolded Conformations of Alpha-lytic Protease Are More Stable Than Its Native State. *Nature*, **395**, 817-819.
- Squitti, R. et al. (2005) Excess of serum copper not related to ceruloplasmin in Alzheimer disease. *Neurology*, **64**, 1040-1046.
- Stahel, P.F., Morganti-Kossmann, M.C. & Kossmann, T. (1998) The role of the complement system in traumatic brain injury. *Brain Research Reviews*, **27**, 243-256.
- Stewart, C.R. et al. (2007) Serum amyloid P co-localizes with apolipoproteins in human atheroma: Functional implications. *J Lipid Res*, **48**, 2162-2171.
- Stoiber, H., Speth, C. & Dierich, M.P. (2003) Role of complement in the control of HIV dynamics and pathogenesis. *Vaccine*, **21**, S77-S82.

- Stover, C. et al. (2001) The human gene for mannan-binding lectin-associated serine protease-2 (MASP-2), the effector component of the lectin route of complement activation, is part of a tightly linked gene cluster on chromosome 1p36.2-3. *Genes Immun*, **2**, 119-127.
- Strang, C.J. et al. (1982) Ultrastructure of the first component of human complement: electron microscopy of the crosslinked complex. *Proc Natl Acad Sci U S A*, **79**, 586-590.
- Tacnet-Delorme, P., Chevallier, S. & Arlaud, G.J. (2001) Beta-amyloid fibrils activate the C1 complex of complement under physiological conditions: evidence for a binding site for A beta on the C1q globular regions. *J Immunol*, **167**, 6374-6381.
- Taylor, J.F., Green, A.A. & Cori, G.T. (1948) Crystalline aldolase. *J Biol Chem*, **173**, 591-604.
- Ten Eyck, L.F. (1973) Crystallographic fast Fourier transforms. *Acta Crystallographica Section A*, **29**, 183-191.
- Thompson, D. et al. (2002) The structures of crystalline complexes of human serum amyloid P component with its carbohydrate ligand, the cyclic pyruvate acetal of galactose. *J Mol Biol*, **320**, 1081-1086.
- Thompson, D., Pepys, M.B. & Wood, S.P. (1999) The physiological structure of human C-reactive protein and its complex with phosphocholine. *Structure*, **7**, 169-177.
- Tillett, W.S. & Francis, T. (1930) Serological reactions in pneumonia with a non-protein somatic fraction of *Pneumococcus*. *J Exp Med*, **52**, 561-571.
- Tissot, B. et al. (2005) Mass Spectrometry Analysis of the Oligomeric C1q Protein Reveals the B Chain as the Target of Trypsin Cleavage and Interaction with Fucoidan. *Biochemistry Biochemistry*, **44**, 2602-2609.
- Tseng, Y., Zavodszky, P. & Schumaker, V.N. (1997) The human complement C1 complex has a picomolar dissociation constant at room temperature. *J Immunol*, **158**, 937-944.
- Uhlenbruck, G. et al. (1978) Additional precipitation reactions of lectins with human serum glycoproteins. *J Clin Chem Clin Biochem*, **16**, 19-23.
- UniProt (2010) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res*, **38**, D142-8.
- Uriu-Adams, J.Y. & Keen, C.L. (2005) Copper, oxidative stress, and human health. *Mol Aspects Med*, **26**, 268-298.
- Villiers, C.L., Arlaud, G.J. & Colomb, M.G. (1985) Domain structure and associated functions of subcomponents C1r and C1s of the first component of human complement. *Proc Natl Acad Sci U S A*, **82**, 4477-4481.

- Wagenaar-Bos, I.G. & Hack, C.E. (2006) Structure and function of C1-inhibitor. *Immunol Allergy Clin North Am*, **26**, 615-632.
- Wagner, E. & Frank, M.M. (2010) Therapeutic potential of complement modulation. *Nat Rev Drug Discov*, **9**, 43-56.
- Walport, M.J. (2002) Complement and systemic lupus erythematosus. *Arthritis Res*, **4**, 279-293.
- Weissman, J.S. (1995) All Roads Lead to Rome? The Multiple Pathways of Protein Folding. *Chem Biol*, **2**, 255-260.
- Wilkins, M.R. et al. (1999) Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol*, **112**, 531-552.
- Windfuhr, J.P., Alsenz, J. & Loos, M. (2005) The critical concentration of C1-esterase inhibitor (C1-INH) in human serum preventing auto-activation of the first component of complement (C1). *Mol Immunol*, **42**, 657-663.
- Wood, S.P. et al. (1988) A pentameric form of human serum amyloid P component. Crystallization, X-ray diffraction and neutron scattering studies. *J Mol Biol*, **202**, 169-173.
- Yeates, T.O. & Fam, B.C. (1999) Protein crystals and their evil twins. *Structure*, **7**, R25-9.
- Yee, A., Gutmanas, A. & Arrowsmith, C.H. (2006) Solution NMR in structural genomics. *Curr Opin Struct Biol*, **16**, 611-617.
- Ying, S.C. et al. (1993) Human serum amyloid P component oligomers bind and activate the classical complement pathway via residues 14-26 and 76-92 of the A chain collagen-like region of C1q. *J Immunol*, **150**, 169-176.
- Yon, J.M. (2002) Protein Folding in the Post-genomic Era. *J Cell Mol Med*, **6**, 307-327.
- Zarkadis, I.K., Mastellos, D. & Lambris, J.D. (2001) Phylogenetic aspects of the complement system. *Dev Comp Immunol*, **25**, 745-762.
- Zhao, H. et al. (2001) Environmentally Induced Reversible Conformational Switching in the Yeast Cell Adhesion Protein Alpha-agglutinin. *Protein Sci*, **10**, 1113-1123.
- Ziccardi, R.J. (1983) The first component of human complement (C1): activation and control. *Springer Semin Immunopathol*, **6**, 213-230.
- Ziccardi, R.J. (1983) Nature of the metal ion requirement for assembly and function of the first component of human complement. *J Biol Chem*, **258**, 6187-6192.
- Ziccardi, R.J. & Cooper, N.R. (1977) The subunit composition and sedimentation properties of human C1. *J Immunol*, **118**, 2047-2052.
- Zipfel, P.F. (2009) Complement and immune defense. *Immunol Lett*, **126**, 1-7.

Appendix I

This appendix contains supplementary data for the biochemical experiments described in the main body of this work, such as calibration data and calculations.

i. C1 charge calculations

Calculations of protein charge were performed using the calculator available at <http://www.scripps.edu/~cdputnam/protcalc.html> (last accessed 8/3/10). At physiological pH, the C1 complex appears to be very close to its predicted isoelectric point ($pI = 7.2$). In this region of the graph (pH 5.5–8.5) the charge curve is approximately linear, and the gradient is shallow, so the predicted net charge of C1 varies little with changes in pH. We can see that at around pH 8.5, the charge curve begins to drop off sharply, in other words C1 becomes more negatively charged more rapidly with increasing pH. Likewise, as the pH drops below approximately pH 5.5, there is a sharp increase in the predicted net charge for a corresponding decrease in pH.

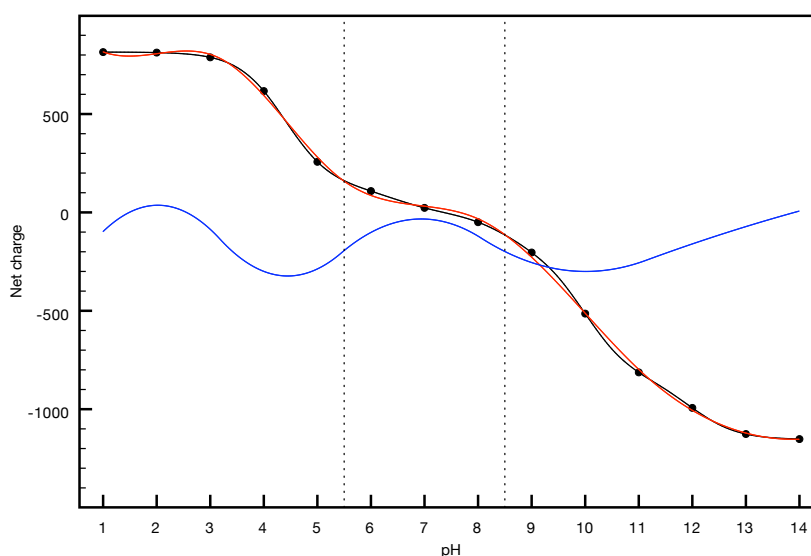


Figure 1.a: Net charge of C1 as a function of pH. The net charge was calculated based on the predicted number of charged residues present in the assembled C1 complex, and the pK_a values of those residues. Calculated values are shown (filled circles), and the data were interpolated by spline interpolation (black). The data were fitted using a spline fitting procedure (red), and the derivative function is also shown (blue). For convenience, pH = 5.5 and pH = 8.5 are also marked with dashed lines. The plot shows that the net charge of C1 is linear and relatively stable in this region, which corresponds to the physiological milieu that C1 is likely to encounter.

ii. Superose 6 calibration

The Superose 6 column was calibrated using a mixture of ribonuclease (14kDa), BSA (67kDa), ferritin (440kDa) and thyroglobulin (669kDa). The void volume of the column was determined using 1mg/ml blue dextran. K_{av} was calculated using the following formula:

$$K_{av} = \frac{V_e - V_o}{V_t - V_o}$$

and plotted against the logarithm of the protein molecular weights. Size exclusion chromatography takes the hydrodynamic radius of molecules into account, so a protein may elute at an apparently higher or lower molecular weight than its true value, and molecular weights determined from this calibration were used as a guideline only.

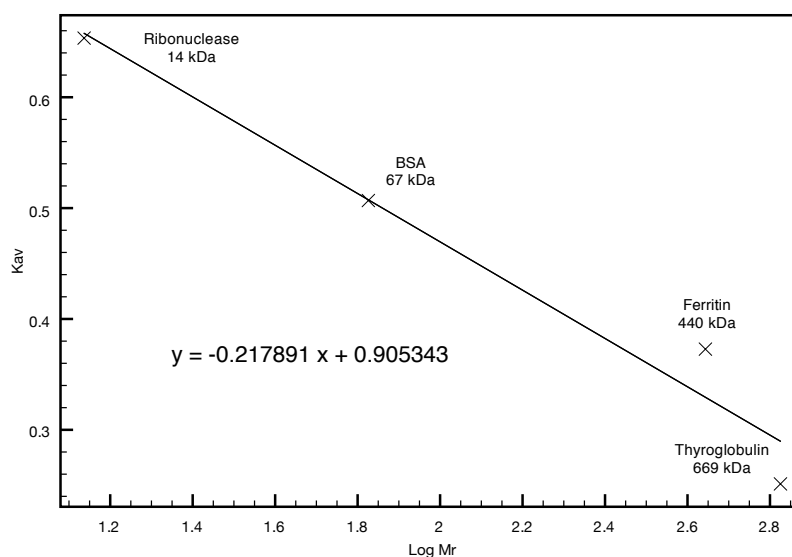


Figure I.b: Superose 6 calibration curve. A linear regression was performed on the data, and the corresponding equation is shown. The calibration curve allows molecular weights of unknown proteins to be estimated within the range of 14–669 kDa.

iii. C1 crystallisation: SEC-purified C1

Fractions containing C1 were pooled and concentrated to 10mg/ml (a 1:10 dilution of the sample gave 0.971mg/ml by Bradford assay against a BSA standard). The concentrated sample was yellowish and opalescent. The protein was screened for precipitation on the addition of small volumes of various common precipitants, and approximate phase diagrams were constructed from the recorded observations. The results of these experiments are presented below. The soluble region is clear, while the region in which the protein precipitates is shaded in red.

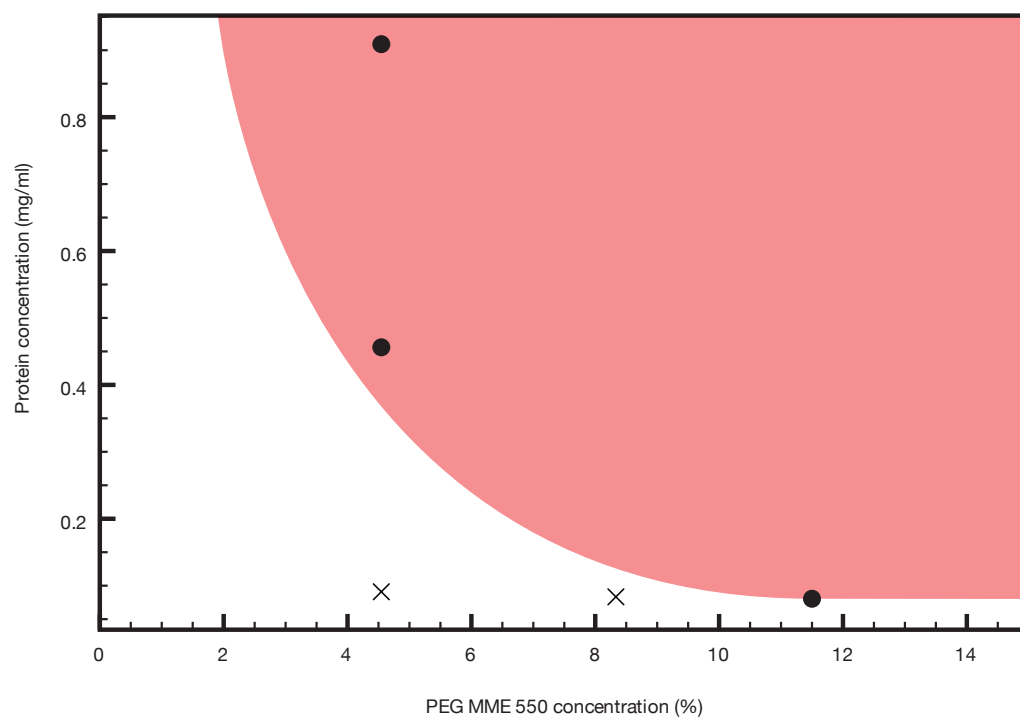


Figure I.c: PEG MME 550 solubility curve.

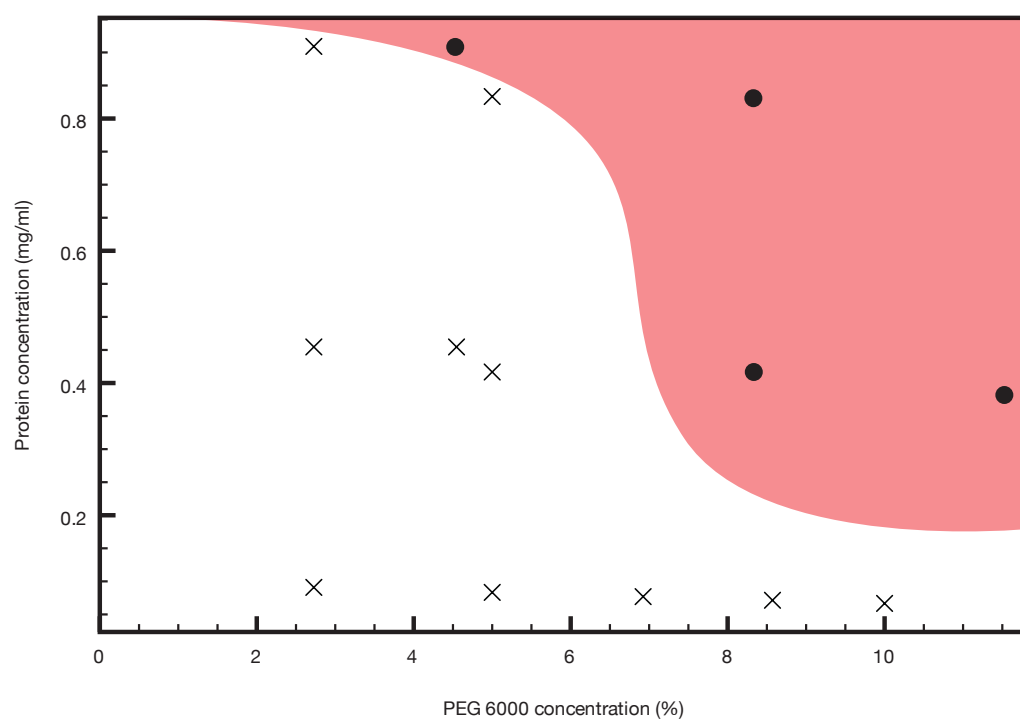


Figure I.d: PEG 6000 solubility curve.

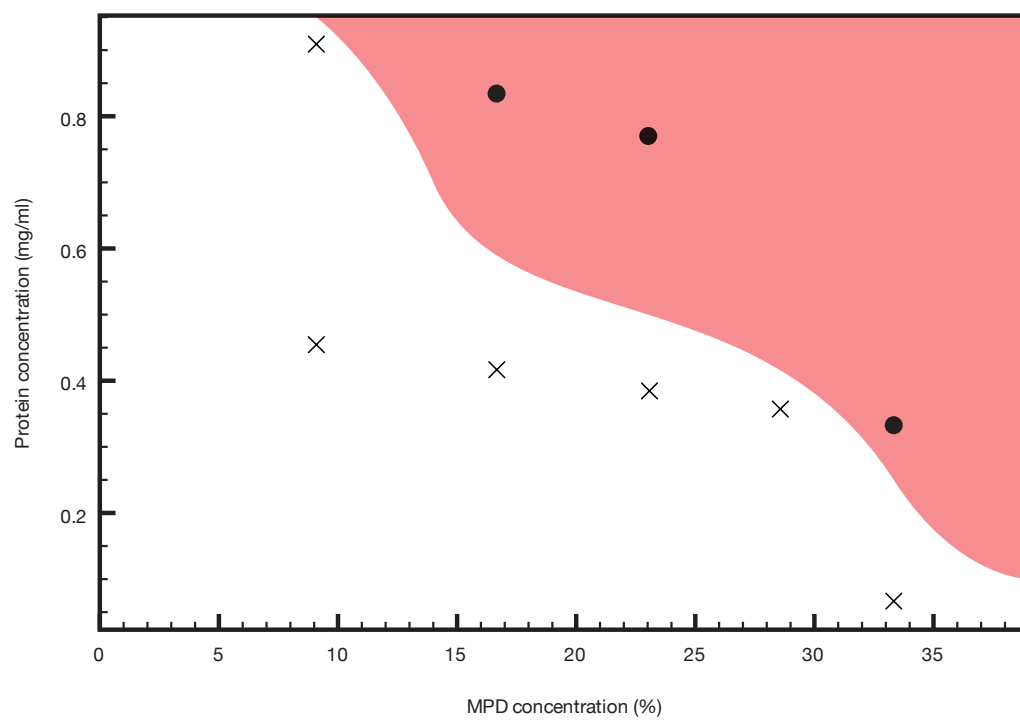


Figure I.e: MPD solubility curve.

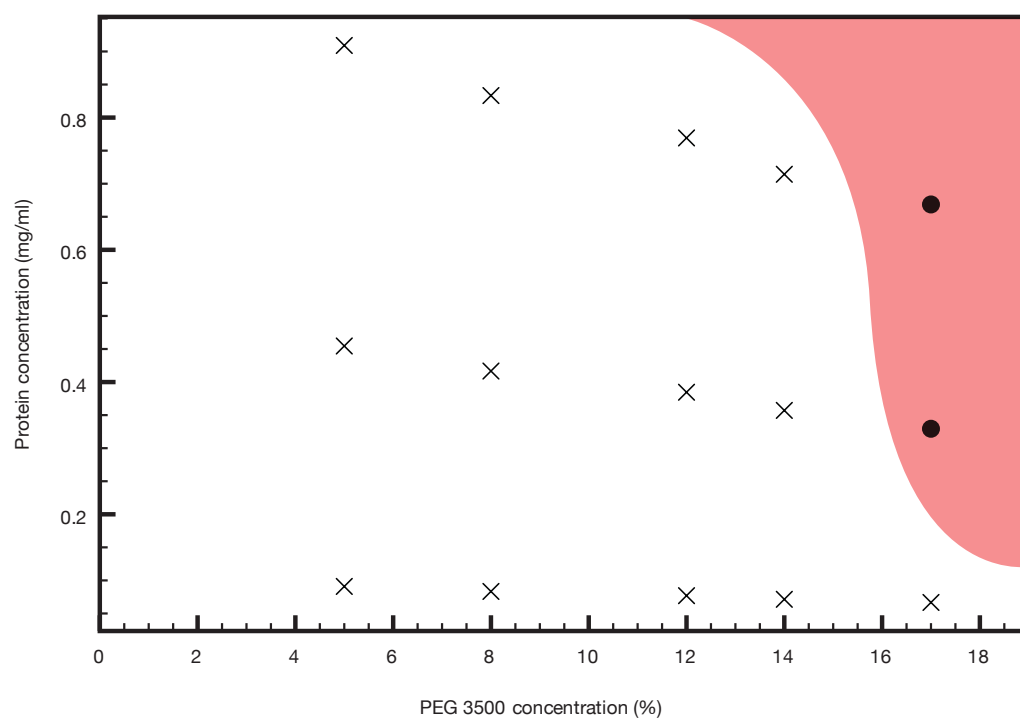


Figure I.f: PEG 3500 solubility curve.B

Appendix II

This appendix contains supplementary data for the crystallographic procedures and calculations described in the main body of this work. The raw diffraction images and relevant calculation files can be found on the DVDs that are included with this thesis.

i. ZiCo1 and ZiCo95 charge calculations

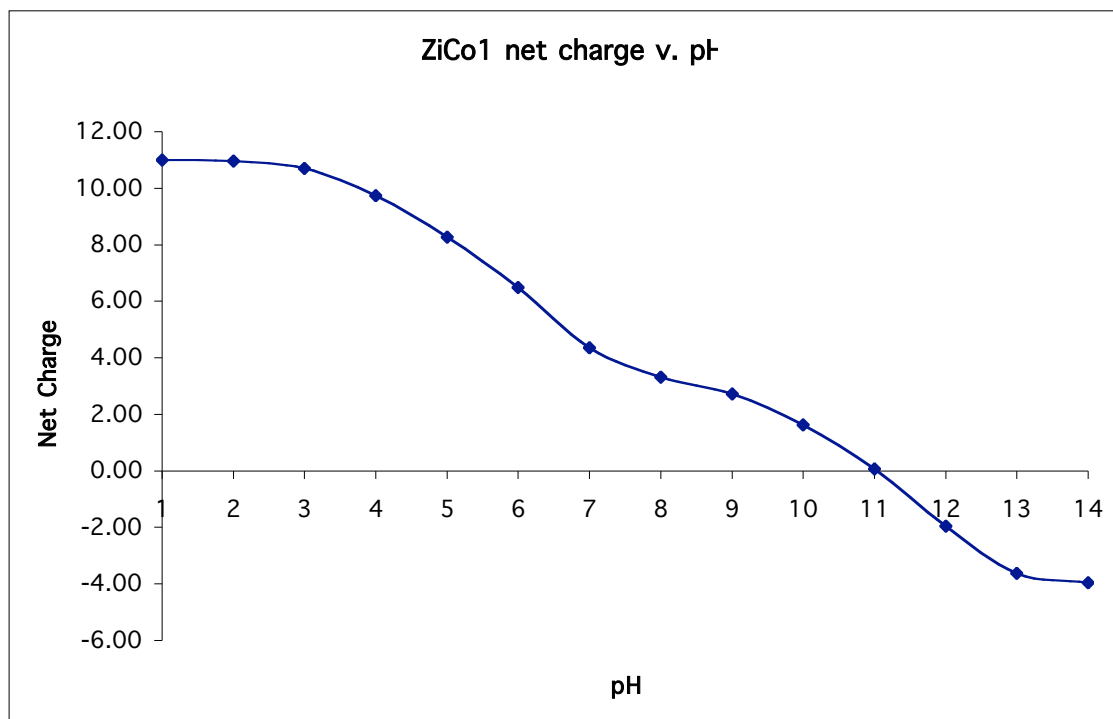


Figure II.a: The net charge of ZiCo1 as a function of pH.

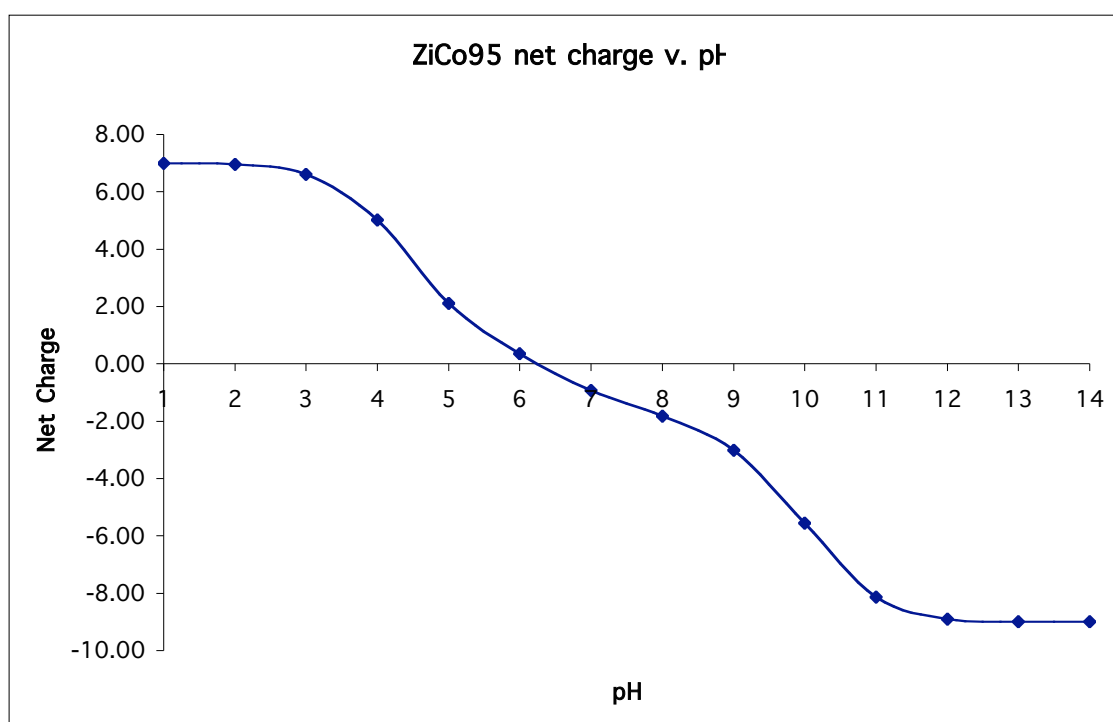


Figure II.b: The net charge of ZiCo95 as a function of pH.

ii. Space groups

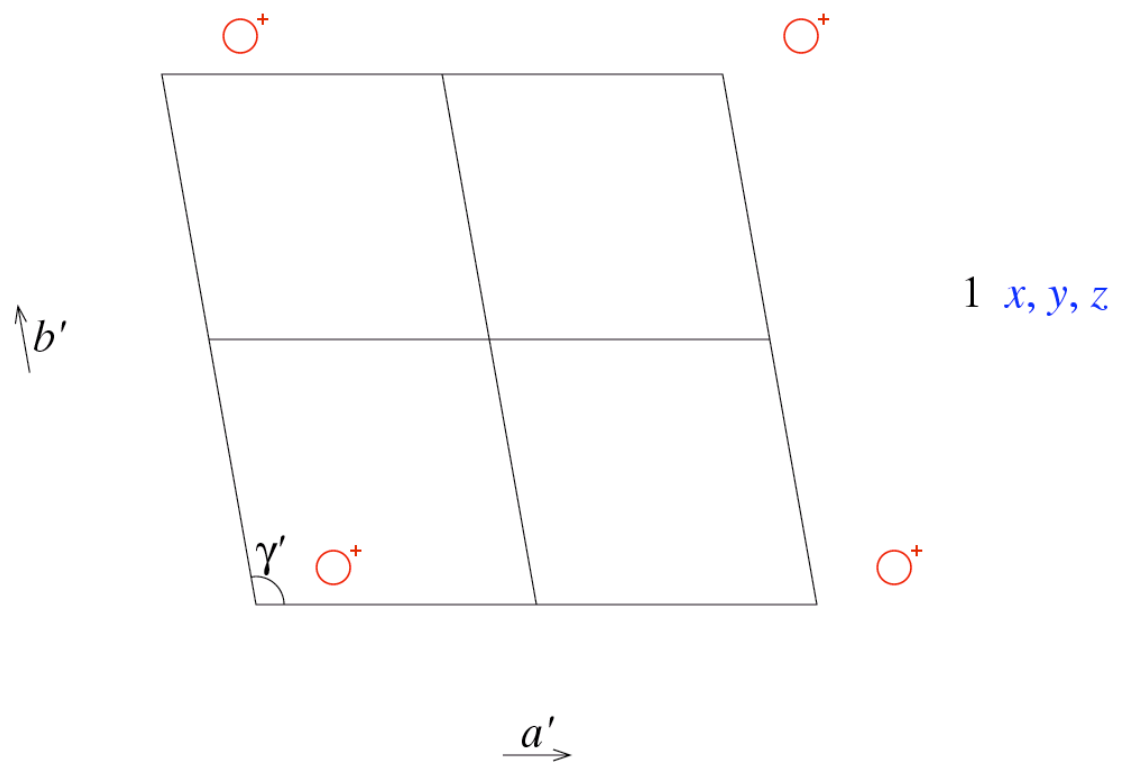


Figure II.c: Space group P1.

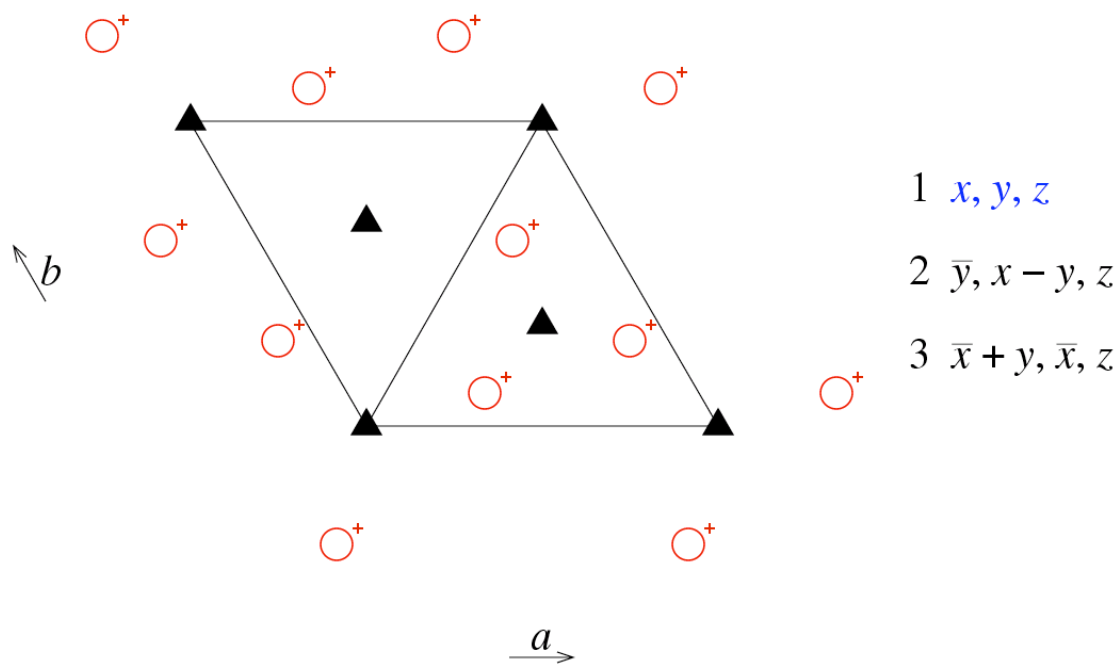


Figure II.d: Space group P3.

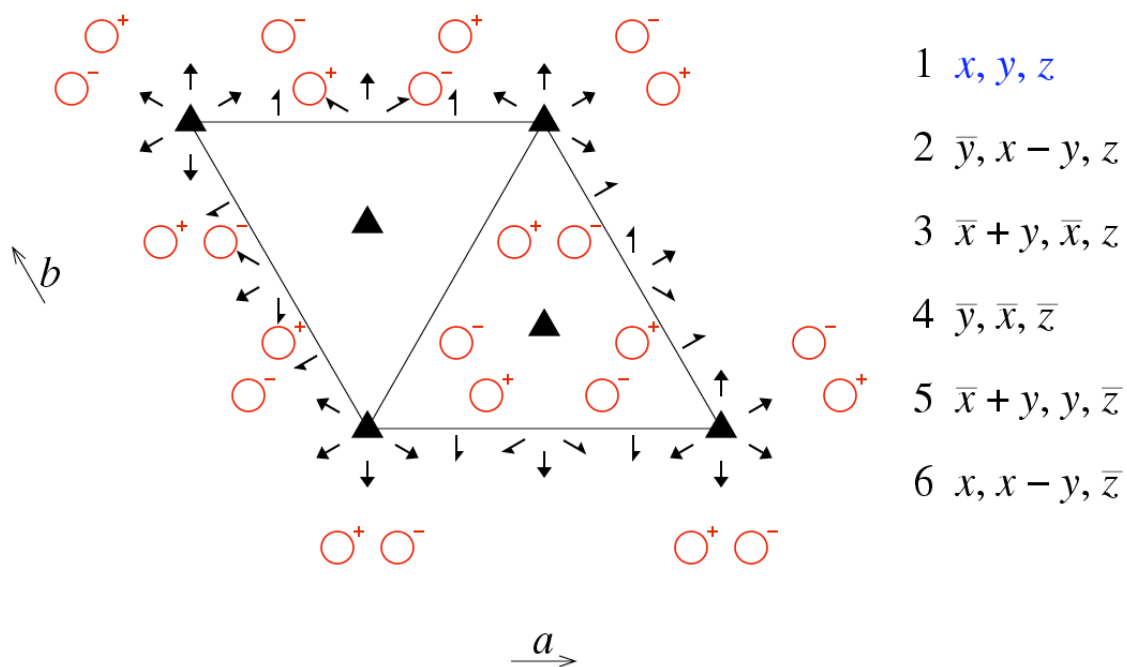


Figure II.e: Space group P312.

iii. Self rotation function

The self rotation function was computed in various space groups using MOLREP and the results are shown in Fig. II.f–Fig. II.k.

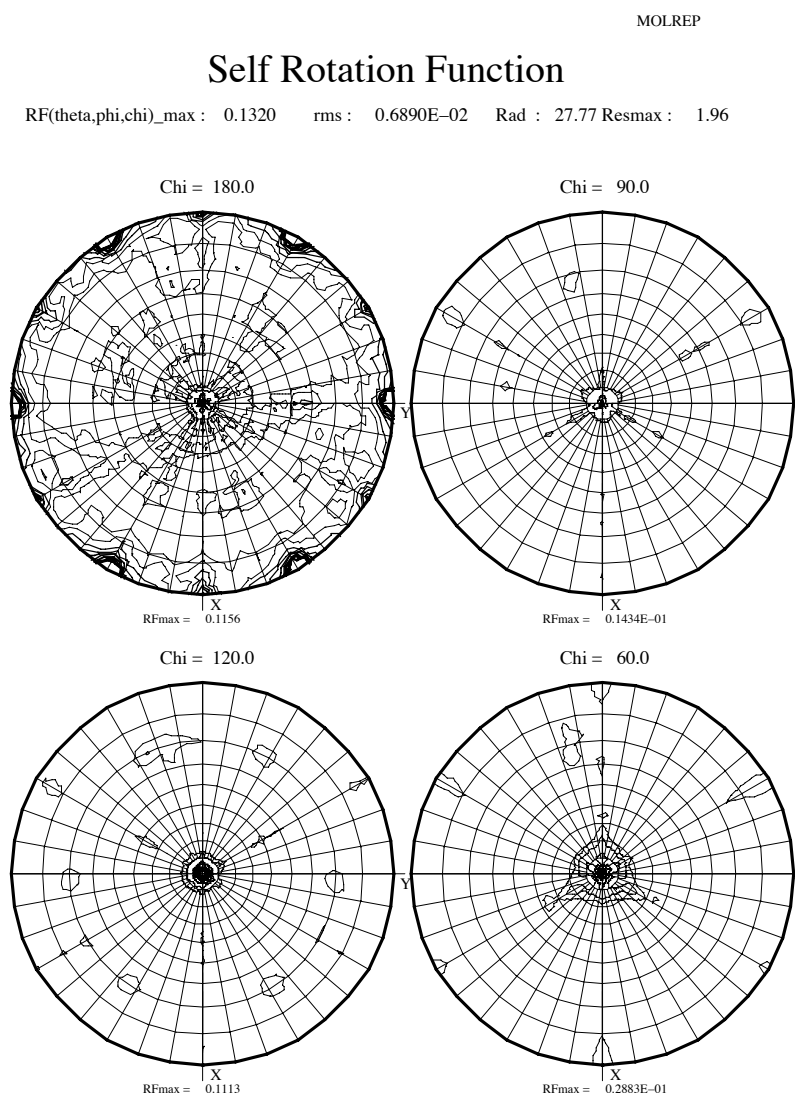


Figure II.f: Stereographic projection of the self rotation function in space group P1. The resolution range (25.06–1.92 Å) and radius of integration (27.77 Å) were determined automatically.

Self Rotation Function

RF(theta,phi,chi)_max : 0.7169E+06 rms : 0.1395E+06 Rad : 7.00 Resmax : 1.97

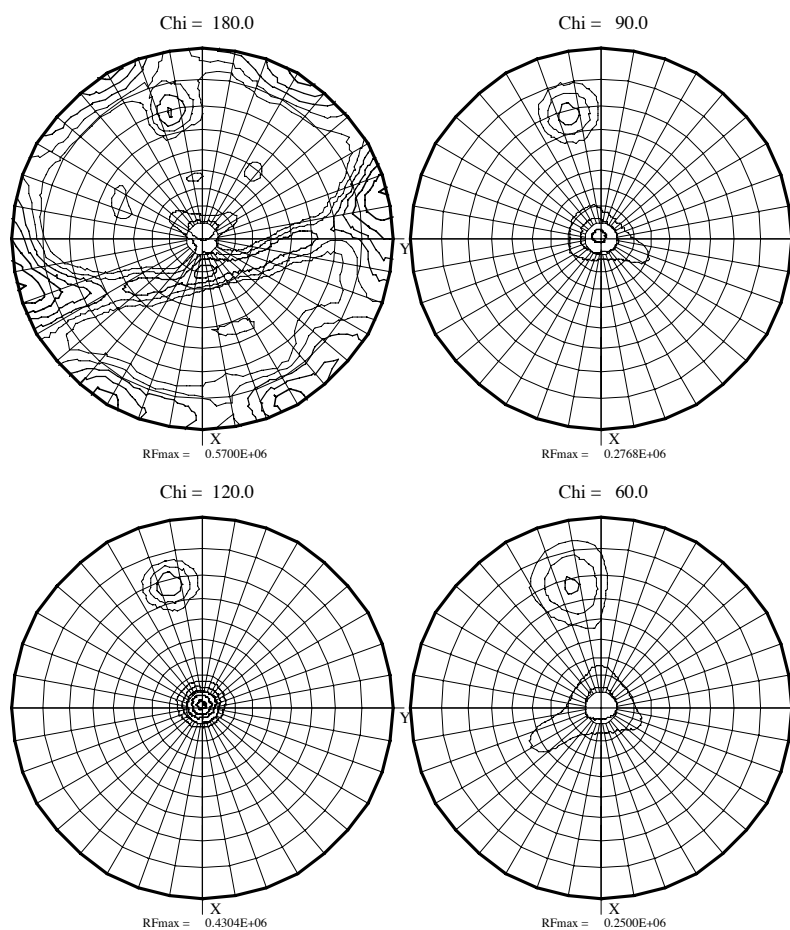


Figure II.g: Modified stereographic projection of the self rotation function in space group P1. The radius of integration was set to 7 Å, which is close to the diameter of a single α -helix, and the resolution range 25.06–1.97 Å) was determined automatically.

Self Rotation Function

RF(theta,phi,chi)_max : 23.25 rms : 1.935 Rad : 19.23 Resmax : 1.86

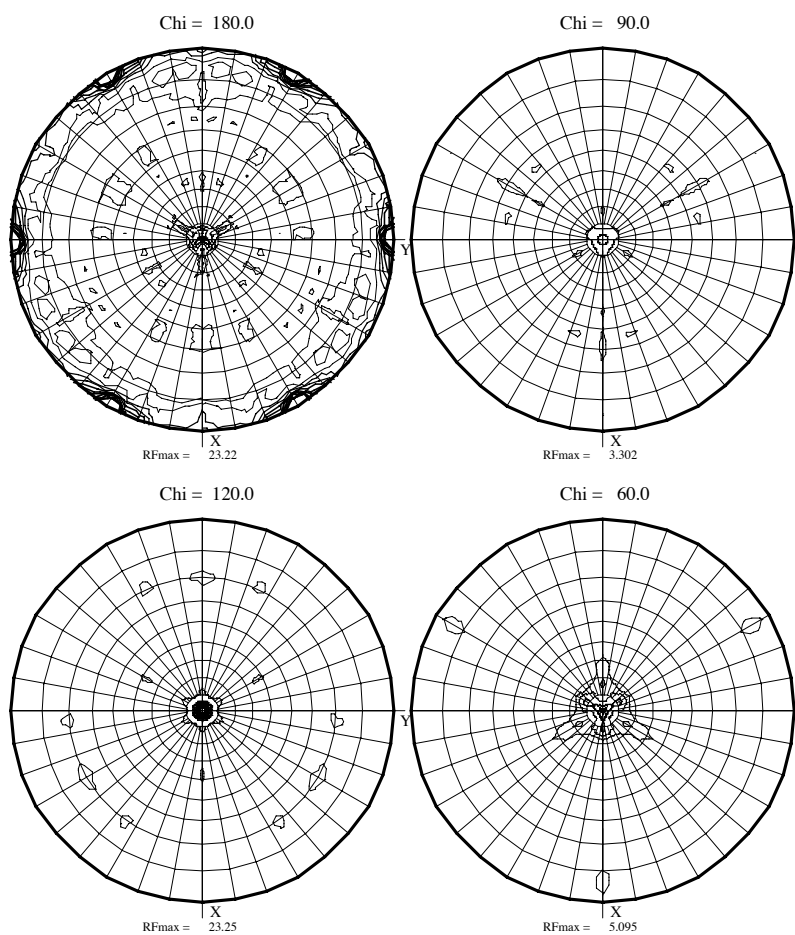


Figure II.h: Stereographic projection of the self rotation function in space group P3. The resolution range (25.06–1.86 Å) and radius of integration (19.23 Å) were determined automatically.

Self Rotation Function

RF(theta,phi,chi)_max : 0.2244E+06 rms : 0.5965E+05 Rad : 7.00 Resmax : 1.86

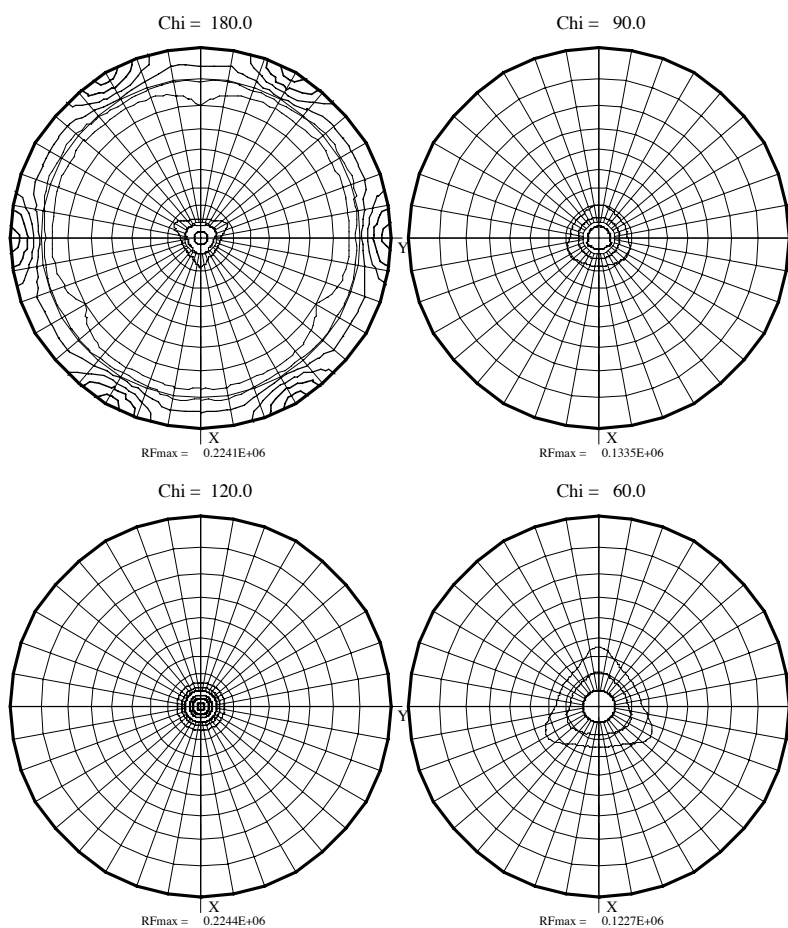


Figure II.i: Modified stereographic projection of the self rotation function in space group P3. The radius of integration was set to 7 Å, which is close to the diameter of a single α -helix, and the resolution range 25.06–1.86 Å) was determined automatically.

Self Rotation Function

RF(theta,phi,chi)_max : 398.4 rms : 40.83 Rad : 15.27 Resmax : 1.84

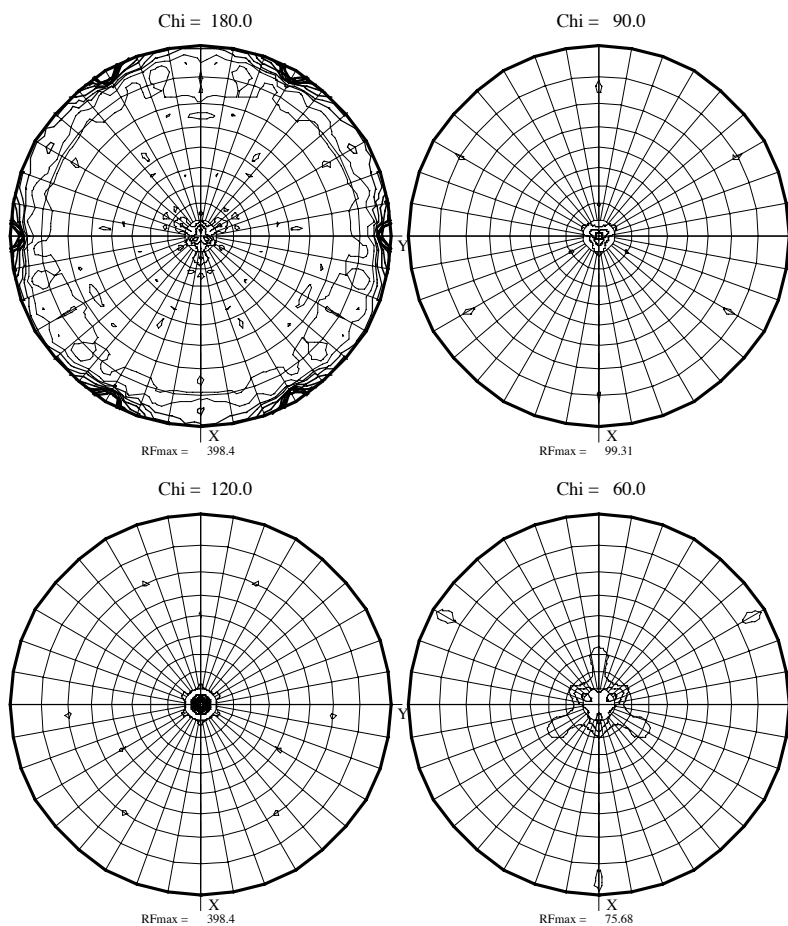


Figure II.j: Stereographic projection of the self rotation function in space group P312. The resolution range (25.06–1.84 Å) and radius of integration (15.27 Å) were determined automatically.

Self Rotation Function

RF(theta,phi,chi)_max : 0.1724E+06 rms : 0.4352E+05 Rad : 7.00 Resmax : 1.92

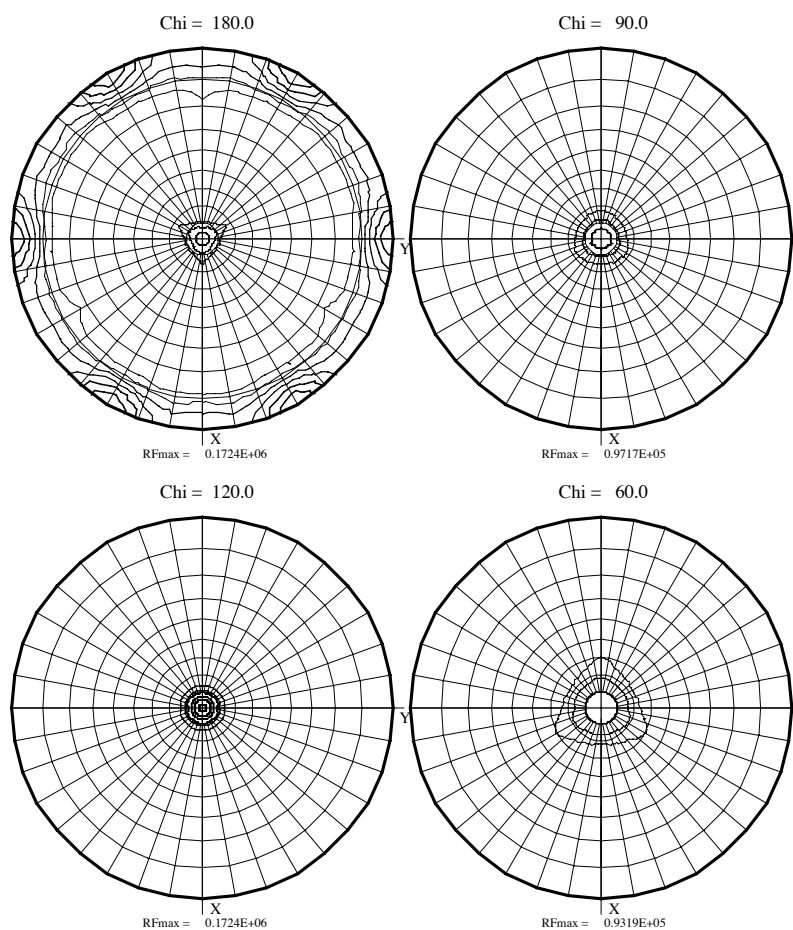


Figure II.k: Modified stereographic projection of the self rotation function in space group P312. The radius of integration was set to 7 Å, which is close to the diameter of a single α -helix, and the resolution range 25.06–1.92 Å) was determined automatically.B

Scio nescio.